STAT 330 (Winter 2013 - 1135)
Mathematical Statistics

Prof. M. Molkaraie
University of Waterloo

LaTeXer: W. Kong
http://wwkong.github.io
Last Revision: April 30, 2014

Table of Contents

These notes are currently a work in progress, and as such may be incomplete or contain errors.

ACKNOWLEDGMENTS:

Special thanks to *Michael Baker* and his LaTeX formatted notes. They were the inspiration for the structure of these notes.

Abstract

The purpose of these notes is to provide a secondary reference for the material covered in STAT 330. Readers should note that nearly 1/3 of the class is spent reviewing concepts learned in STAT 230/231 but later material can prove to be significantly more difficult. The author recommends that students who enroll in this should have a very good background in calculus as that is the core of the computations done in this course.

Overview

1) Review	2) Joint, Marginal, and Conditional Distributions
3) Functions of RVs	4) Convergence in P & in D
5) Point Estimation and Maximum Likelihood	6) Hypothesis Testing

Recommended Readings:

- Intro to Probability and Mathematical Statistics (Bain & Englehardt)

- Statistical Inference (Casella & Berger)

Test 1 on June 12th

Test 2 on July 17th

1 Review

We briefly go over some basic concepts introduced in STAT 230 and STAT 231

1.1 Probability Spaces

Definition 1.1. Recall that a *probability space* is composed of a set S, called the sample space or the set of all possible outcomes (also sometimes given by Ω) where $E \subset S$ is called an event, a sigma algebra Σ, generated by S, and a probability function $P : \Sigma \mapsto \mathbb{R}^n$ where n is usually 1.

Axiom 1. *Here are the properties of the probability function (Kolmogorov Axioms):*

1) $P(A) \geq 0, \forall A \subset S$

2) $P(S) = 1$

3) If $A = \bigcup\limits_{i \in I} A_i$ *are disjoint and* I *is countable then* $P(A) = \sum\limits_{i \in I} P(A_i)$

Note that 3) is also known as $\sigma-$*additivity.*

Definition 1.2. We define the conditional probability $P(A|B) = \frac{P(A \cap B)}{P(B)}$ where $P(B) \neq 0$, and we say that A and B are independent or $A \perp B$ if $P(A) = P(A|B) \implies P(A \cap B) = P(A)P(B)$.

Note 1. Independence $\not\Leftrightarrow$ Disjoint

1.2 Random Variables

Definition 1.3. A random variable is a function $X : S \mapsto \mathbb{R}$ with the following properties and notations:

$$\begin{aligned} \{X \leq x\} &= \{w : w \in S, X(w) \leq x\} \\ P\{X \leq x\} &= P\{w : w \in S, X(w) \leq x\} = F_X(x) \end{aligned}$$

where the second form $F_X : \mathbb{R}_{EXT} \mapsto [0, 1]$, is known as the cdf or *cumulative distribution function*.

Proposition 1.1. *The cdf has the following properties:*

0) F is non-decreasing

1) $F_X(-\infty) = 0$

2) $F_X(\infty) = 1$

2.5) $x_1 < x_2 \implies F_X(x_1) \leq F_X(x_2)$

3) F_X is right continuous

4) $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$, for $a < b$

5) $P(X = b) = F_X(b) - \lim\limits_{x \to b^-} F_X(x)$, equal to 0 if X is continuous

Example 1.1. Consider $T(x) = \frac{1}{1-e^{-x}}$, $x \in \mathbb{R}$. By observation it satisfies 1), 2), and 3). To check 0), note that

$$\frac{dT(x)}{dx} = \frac{e^{-x}}{(1+e^{-x})^2} > 0$$

Definition 1.4. For discrete random variables, say X, in addition to a cdf, we define a *probability mass function*, called a pmf:

$$f_X(x) = P_X(X = x) = P_X(x)$$

Proposition 1.2. *Here are some properties of the pmf:*

1) $f_X(x) = P(X = x) \geq 1$

2) $\sum\limits_k P(X = k) = 1$

Example 1.2. Here is a small list of some discrete distributions: Uniform $[Unif(a, b)]$, Geometric $[Geo(p)]$, Poisson $[Poisson(\lambda)]$, Binomial $[Bin(n, p)]$

Example 1.3. Suppose we have a red balls and b black balls

a) Let $X_1 =$ the # of red balls in n selections without replacement

The PMF is Hyper-Geometric $\implies P(X = x) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}$

b) Let $X_1 =$ the # of red balls in n selections with replacement

The PMF is $Bin\left(n, \frac{a}{a+b}\right)$.

Aside. $\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k}\binom{n}{n-k} \implies \binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}^2$ (Called the Vandermonde identity)

Example 1.4. Suppose $f_X(x) = \frac{-(1-p)}{x \ln p}$. We will show that $\sum\limits_{x \in \mathbb{N}} f_X(x) = 1$ (it is clearly non-negative). Observe that

$$\sum_{x=1}^{\infty} \frac{-(1-p)^x}{x \ln p} = \frac{-1}{\ln p} \sum_{x=1}^{\infty} \frac{(1-p)^x}{x}$$

and since

$$-\ln(1-x) = \sum_{n=1}^{\infty} \frac{x^n}{n}, |x| < 1$$

we have

$$\sum_{x=1}^{\infty} \frac{-(1-p)^x}{x \ln p} = \frac{-1}{\ln p} \times (-\ln p) = 1$$

Definition 1.5. For continuous random variables, instead of a pmf $P_X(X = k)$ we have a pdf $f_X(x)$ called a probability density function. Continuous random variables still have a cdf, denoted by $F_X(x) = P(X \leq x)$. Continuous random variables also have the following properties:

1) $f_X(x) \geq 0$

2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

3) $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$

Definition 1.6. The uniform distribution $X \sim \text{Uniform}([a, b])$ has cdf

$$F_X(x) = \int_a^x \frac{1}{b-a} dx = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & x > b \end{cases}$$

Example 1.5. Consider the pdf

$$f_X(x) = \begin{cases} \frac{\theta}{x^{\theta+1}} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

This is a valid pdf provided that

$$\begin{aligned} \int_1^{\infty} f_X(x) = -x^{-\theta}\Big|_1^{\infty} = 1 \quad &\implies \quad 1^{-\theta} - (\infty)^{-\theta} \\ &\implies \quad \theta > 0 \end{aligned}$$

1.3 The Gamma Function

Definition 1.7. The gamma function $\Gamma(\alpha)$ is defined by

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy = \int_0^{\infty} y^{\alpha} \frac{e^{-y}}{y} dy$$

and has the following properties.

1) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$

2) $\Gamma(n) = (n-1)!, n \in \mathbb{N}$

3) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Definition 1.8. The gamma distribution $X \sim \text{Gamma}(\alpha, \beta)$ is defined using the pdf

$$f_X(x) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}, x \geq 0, \alpha > 0, \beta > 0$$

and note that

$$\begin{aligned} I &= \int_0^\infty \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^{\alpha-1}e^{-\frac{x}{\beta}}dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty \beta^{\alpha-1}y^{\alpha-1}e^{-y}\beta dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty y^{\alpha-1}e^{-y}dy \\ &= \frac{\Gamma(\alpha)}{\Gamma(\alpha)} \\ &= 1 \end{aligned}$$

with $\frac{x}{\beta} = y \implies dx = \beta dy \implies \beta y = x$. So the gamma distribution is a valid distribution.

Definition 1.9. The Weibull distribution $X \sim$ Weibull(θ, β) is given by the pdf

$$f_X(x) = \frac{\beta}{\theta^\beta}x^{\beta-1}e^{-\left(\frac{x}{\theta}\right)^\beta} = \frac{\beta}{\theta^\beta}x^{\beta-1}exp\left(-\left(\frac{x}{\theta}\right)^\beta\right), x > 0, \theta > 0, \beta > 0$$

and note that for $\beta = 1$, we have a Exponential(θ) distribution. To see that it is a valid pdf, observe that

$$\begin{aligned} I &= \int_0^\infty f_X(x)dx \\ &= \int_0^\infty \frac{\beta}{\theta^\beta}x^{\beta-1}e^{-\left(\frac{x}{\theta}\right)^\beta}dx \\ &= \int_0^\infty e^{-y}dy = 1 \end{aligned}$$

using $y = \left(\frac{x}{\theta}\right)^\beta \implies dy = \frac{1}{\theta^\beta}\beta x^{\beta-1}dx$. Suppose that $\theta = 1$. Then

$$f_X(x) = \beta x^{\beta-1}e^{-x^\beta}, F_X(x) = 1 - e^{-x^\beta}, x \geq 0$$

2 Expectation and Variance

We briefly go over the definitions and properties of *expectation* and *variance*.

2.1 Expectation

Definition 2.1. The expectation of a random variable X denoted as $E[X], E(X), EX, \mu_X$ or μ is defined as

$$E(X) = \sum_{x\in\mathbb{Z}} xP(X = x)$$

in the discrete case and

$$E(X) = \int_{-\infty}^\infty xf_X(x)dx$$

for the continuous case. We will illustrate examples and properties of expectation in the continuous case from this point forward. For general functions of random variables, $g(X)$, we have

$$E(g(X)) = \int_{-\infty}^\infty g(x)f_X(x)dx$$

and for joint expectations $E(XY)$ we have

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)dxdy$$

Summary 1. Some properties of expectation are as follows.

1) Linearity of Expectation: $E(a \cdot g(X) + b \cdot g(Y)) = aE(X) + bE(Y)$ even if X is dependent on Y

2) $X \perp Y \implies E(XY) = E(X)E(Y)$

2.2 Variance

Definition 2.2. We define the variance of a random variable X as

$$E[(X - E[X])^2] = E[X^2] - E^2[X] = EX^2 - \mu_X^2$$

which usually denote as $Var(X), \sigma_X$ or σ. Note that $EX^2 \geq (EX)^2$ and these are equal when X is a constant.

Definition 2.3. We define the following moments around X

1) k^{th} moment: $E[X^k]$

2) k^{th} moment around the mean (central moment): $E\left[(X-\mu)^k\right]$

3) k^{th} factorial moment: $E\left[x^{(k)}\right] = E[X(X-1)...(X-k+1)]$

Example 2.1. Suppose that $X \sim N(0,1)$, then $EX^{2k+1} = 0, \forall k \in \mathbb{N}$ since the integrand is the product of a symmetric (even) and antisymmetric (odd) function.

Summary 2. Here are some properties of the variance function.

1) $Var(aX + b) = a^2 Var(X)$

2) $Var(aX + bY) = a^2 Var(X) + b^2 Var(X) + 2abCov(X,Y)$

Example 2.2. If $X \sim Pois(\theta)$ then $E\left[X^{(k)}\right] = \theta^k$. To see this, we use the definition below.

$$\begin{aligned}
E\left[x^{(k)}\right] &= E[X(X-1)...(X-k+1)] \\
&= e^{-\theta}\sum_{0\leq x<\infty}\frac{x^{(k)}\theta^k}{x!} \\
&= e^{-\theta}\sum_{k\leq x<\infty}\frac{x^{(k)}\theta^k}{x!} \\
&= e^{-\theta}\sum_{1\leq x<\infty}\frac{\theta^k}{(x-k)(x-k-1)...1} \\
&= e^{-\theta}\sum_{0\leq y<\infty}\frac{\theta^k}{y!}, y = x-k \\
&= \theta^k e^{-\theta}\sum_{0\leq y<\infty}\frac{1}{y!} \\
&= \theta^k
\end{aligned}$$

using this result, we can deduce that

$$EX = \theta, E(X(X-1)) = \theta^2 = E(X^2) - E(X) \implies Var(X) = \theta$$

Example 2.3. If $X \sim \text{Gamma}(\alpha, \beta)$ then $E[X^p] = \beta^p \left(\frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}\right)$ and to see this, we use the definition again

$$\begin{aligned}
E[X^p] &= \int_{-\infty}^{\infty} x^p \frac{x^{\alpha-1} e^{\frac{-x}{\beta}}}{\Gamma(\alpha)\beta^\alpha} dx \\
&= \int_{-\infty}^{\infty} \frac{x^{\alpha+p-1} e^{\frac{-x}{\beta}}}{\Gamma(\alpha)\beta^\alpha} dx \\
&= \int_{-\infty}^{\infty} \frac{y^{\alpha+p-1}\beta^{\alpha+p-1}}{\Gamma(\alpha)\beta^\alpha} e^{-y}\beta dy, y = \frac{x}{\beta} \\
&= \int_{-\infty}^{\infty} \frac{y^{\alpha+p-1}\beta^p}{\Gamma(\alpha)} e^{-y} dy \\
&= \frac{\beta^p}{\Gamma(\alpha)} \int_{-\infty}^{\infty} y^{\alpha+p-1} e^{-y} dy \\
&= \beta^p \left(\frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}\right)
\end{aligned}$$

We then use this to get $E[X] = \beta\alpha$ and $E[X^2] = \beta^2\alpha(\alpha-1)$ with $Var(X) = \beta^2\alpha$. Note that if we know $E[X]$ and $Var[X]$, we can solve for α and β.

3 Moment Generating Functions (MGFs)

Definition 3.1. A *moment generating function* of X is created by the following mapping

$$(X \mapsto M_X(t)) = E\left[e^{tX}\right] = M_X(t)$$

Example 3.1. Let $X \sim Bin(n,p)$. Then

$$E\left[e^{tX}\right] = \sum_{0 \le x \le n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{0 \le x \le n} \binom{n}{x} \left(e^t p\right)^x (1-p)^{n-x} = \left[pe^t + (1-p)\right]^n$$

Example 3.2. If $X \sim Pois(\theta)$ then $M_X(t) = e^{\theta(e^t-1)}$. To see this, we go by definition.

$$E\left[e^{tX}\right] = \sum_{0 \le x < \infty} e^{tx} \frac{e^{-\theta}\theta^x}{x!} = e^{-\theta} \sum_{0 \le x < \infty} \frac{\left(e^t\theta\right)^x}{x!} = e^{-\theta} e^{\theta e^t} = e^{\theta(e^t-1)}$$

As a side remark note that if $X \sim Bin(n,p)$ and $n \to \infty, p \to 0$ with $np \to \theta$ then $X \to Pois(\theta)$. We also should get that their moment generating functions should converge.

Example 3.3. If $X \sim \text{Gamma}(\alpha, \beta)$ then $M_X(t) = (1-\beta t)^{-\alpha}$. To see this, we also go by definition

$$\begin{aligned}
M_X(t) &= E\left[e^{tX}\right] \\
&= \int_0^{\infty} e^{tx} \frac{x^{\alpha-1} e^{\frac{-x}{\beta}}}{\Gamma(\alpha)\beta^\alpha} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^{\infty} x^{\alpha-1} e^{-x\left(\frac{1}{\beta}-t\right)} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^{\infty} \frac{y^{\alpha-1}}{\left(\frac{1}{\beta}-t\right)^{\alpha-1}} \cdot e^{-y} \frac{dy}{\left(\frac{1}{\beta}-t\right)}, y = \left(\frac{1}{\beta}-t\right) x \\
&= \beta^{-\alpha} \left(\frac{1}{\beta}-t\right)^{-\alpha} \int_0^{\infty} \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy \\
&= \beta^{-\alpha} \left(\frac{1}{\beta}-t\right)^{-\alpha} = (1-\beta t)^{-\alpha}
\end{aligned}$$

Example 3.4. Suppose that $Z \sim N(0,1)$. Then $M_Z(t) = e^{\frac{t^2}{2}}$. As above, we go by definition.

$$M_Z(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} e^{\frac{t^2}{2}} dx = e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} dx = e^{\frac{t^2}{2}}$$

3.1 Linear Combinations

Proposition 3.1. *Given the MGF of X, we can compute the MGF of any linear combination of X, say $Y = aX + b$.*

Proof. We can do this directly.

$$M_Y(t) = E\left[e^{Yt}\right] = E\left[e^{(aX+b)t}\right] = e^{bt} M_X(at)$$

□

Corollary 3.1. *If $Y \sim N(\mu, \sigma^2)$, what is $M_Y(t)$? Well if $X \sim N(0,1)$, then $Y = \mu + \sigma X$. Hence*

$$M_Y(t) = e^{\mu t} e^{\frac{\sigma^2 t^2}{2}} = e^{\frac{t(2\mu + \sigma^2 t)}{2}}$$

Summary 3. Recall that

$$e^{tX} = \sum_{k=0}^{\infty} \frac{(tx)^k}{k!} \implies E\left[e^{tX}\right] = \sum_{k=0}^{\infty} \frac{t^k}{k!} E\left[x^k\right]$$

Here are some properties of the MGF:

1) $M_X(0) = 1$

2) $M'_X(t) = \sum_{k=0}^{\infty} \frac{kt^{k-1}}{k!} E\left[x^k\right] \implies M'_X(0) = E[X]$

3) $M''_X(t) = \sum_{k=0}^{\infty} \frac{k(k-1)t^{k-2}}{k!} E\left[x^k\right] \implies M''_X(0) = E[X^2]$

4) Inductively, we can get $M_X^{(n)}(0) = E[X^n]$

5) $M_X = M_Y \implies F_X = F_Y$ (only in this course; generally this is not true)

6) If $Y = \sum_{i=1}^n X_i$, then $M_Y(t) = \prod_{i=1}^n M_{X_i}(t)$

Example 3.5. If $X \sim$ Gamma(α, β) then

$$M'_X(t) = \alpha\beta(1 - \beta t)^{-\alpha-1} \implies E[X] = M'_X(0) = \alpha\beta$$

$$M''_X(t) = \alpha\beta(\alpha+1)\beta(1-\beta t)^{-\alpha-2} \implies E[X^2] = M''_X(0) = \alpha\beta^2(\alpha+1)$$

$$Var[X] = \alpha\beta^2(\alpha+1) - \alpha^2\beta^2 = \alpha\beta^2$$

Example 3.6. Suppose that $M_X(t)$. Find the MGF of $Y = 2X - 1$ and $E[Y], Var[Y]$. What is the distribution of Y? By observation,

$$M_Y(t) = e^{-t} e^{\frac{4t^2}{2}}, E[Y] = -1, Var[Y] = 4$$

and the distribution of Y is $N(-1, 4)$.

3.2 Characteristic Function

Definition 3.2. The *characteristic function* of a random variable X is the Fourier transform of the pdf/pmf:

$$\Im_X^{(\omega)} = \int_x e^{-i\omega x} f_X(x) dx, \|e^{-i\omega x}\| = 1$$

where it always exists and has all of the properties of the MGF.

4 Joint Distributions

Example 4.1. Consider rolling two dice, D_1 and D_2. Let $X = D_1 + D_2$ and $Y = |D_1 - D_2|$. Then

$$P_{XY}(X = 5, Y = 3) = P(X = 5, Y = 3) = \frac{2}{36} = \frac{1}{18}$$

and

$$P(X = 7, Y \leq 4) = \sum_{y=1}^{4} P(X = 7, Y = y) = \frac{4}{36} = \frac{1}{9}$$

Summary 4. Here are some basic properties of the marginals of a joint distribution in the discrete case:

1) $F_1 = F_X(x) = P(X \leq x) = \lim_{y\to\infty} F_{xy}(x, y) = F_{xy}(x, \infty)$

2) $\sum_x \sum_y f_{xy}(x, y) = 1$

3) $P_X(X = x) = \sum_y P(X = x, Y = y)$, $P_Y(Y = y) = \sum_x P(X = x, Y = y)$

and now in the continuous case:

1) $f_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y)$

2) $F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(s, t)\, ds\, dt$

3) $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(s, t)\, ds$, $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(s, t)\, dt$

Example 4.2. Suppose that we have 10 ActSc students, 9 Stats student and 6 Maths students. We select 5 students without replacement. Let $X = \#$ of ActSc students and $Y = \#$ of Stats students.

The joint PMF of X and Y is

$$P_{XY}(X = x, Y = y) = \frac{\binom{10}{x}\binom{9}{y}\binom{6}{5-x-y}}{\binom{25}{5}}, 0 \leq x, y, x + y \leq 5$$

the marginal of X is

$$P_X(X = x) = \frac{\sum_y \binom{10}{x}\binom{9}{y}\binom{6}{5-x-y}}{\binom{25}{5}} = \frac{\binom{10}{x}}{\binom{25}{5}} \sum_y \binom{9}{y}\binom{6}{5-x-y} = \frac{\binom{10}{x}\binom{15}{5-x}}{\binom{25}{5}}$$

and similarly the marginal of Y is

$$P_Y(Y = y) = \frac{\sum_x \binom{10}{x}\binom{9}{y}\binom{6}{5-x-y}}{\binom{25}{5}} = \frac{\binom{9}{y}}{\binom{25}{5}} \sum_x \binom{10}{x}\binom{6}{5-x-y} = \frac{\binom{9}{y}\binom{16}{5-y}}{\binom{25}{5}}$$

Example 4.3. Let

$$f_{XY} = \begin{cases} x+y & 0 \le x, y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Let's check if $\int\int f_{XY}\, d(x,y) = 1$.

$$I = \int_0^1 \int_0^1 (x+y)\, dx\, dy = \int_{y=0}^{y=1} \left(\frac{x^2}{2} + xy\right)\Big|_0^1 dy = \int_{y=0}^{y=1} \left(y + \frac{1}{2}\right) dy = \frac{1}{2} + \frac{1}{2} = 1$$

Let's try to compute $P(X \le \frac{1}{3}, Y \le \frac{1}{2})$.

$$I = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{3}} (x+y)\, dx\, dy = \int_{y=0}^{y=\frac{1}{2}} \left(\frac{x^2}{2} + xy\right)\Big|_0^{\frac{1}{3}} dy = \int_{y=0}^{y=\frac{1}{2}} \left(\frac{1}{3}y + \frac{1}{18}\right) dy = \frac{1}{36} + \frac{1}{24} = \frac{5}{72}$$

Note that the cdf of this pdf is

$$F_{X,Y}(x,y) = \begin{cases} 0 & x, y \le 0 \\ \frac{xy}{2}(x+y) & 0 \le x, y \le 1 \\ 1 & x, y \ge 1 \end{cases}$$

(In the lecture here, we reviewed how to integrate over arbitrary regions so I will only give the important details)

Summary 5. (1) If we are asked to compute $P(f(X,Y) < c)$ for some constant c and random variables (r.v.s) X and Y, isolate Y, draw the region of integration and derive the appropriate integrals. For example, if $0 \le x, y \le 1$, then

$$Pr\left(X + Y < \frac{1}{2}\right) = Pr\left(Y < \frac{1}{2} - X\right) = \int_{x=0}^{\frac{1}{2}} \int_{y=0}^{\frac{1}{2}-x} f_{XY}(x,y)\, dy\, dx$$

and

$$Pr\left(XY \le \frac{1}{2}\right) = Pr\left(Y \le \frac{1}{2X}\right) = 1 - P\left(Y > \frac{1}{2X}\right) = 1 - \int_{x=\frac{1}{2}}^{1} \int_{y=\frac{1}{2x}}^{1} f_{XY}(x,y)\, dy\, dx$$

where in the second example, $x = 0 \implies y = \frac{1}{2}$ and $y = 0 \implies x = \frac{1}{2}$.

(2) If $X \perp Y$ then

$$\iint_A f_{XY}(x,y) d(x,y) = \iint_A f_X(x) f_Y(y) d(x,y)$$

Exercise 4.1. Given $f_{XY} = ke^{-y-y}$, and $0 < x < y < \infty$,

(0) What is k? (Ans: $k = 2$)

(1) What is $P(X \le \frac{1}{3}, Y \le \frac{1}{3})$? (Ans: $1 - e^{-\frac{2}{3}} - 2e^{-\frac{1}{2}}(1 - e^{-\frac{1}{3}})$)

(2) What is $P(X < Y)$? (Ans: 1)

(3) What is $P(X + Y \ge 1)$? (Hint: $P(X + Y \ge 1) = 1 - P(X + Y < 1)$, Ans: $2e^{-1}$)

(4) Are X and Y independent? (Ans: No! Check the marginals.)

Definition 4.1. We define the *support* of a r.v. as $\{x : f_X(x) > 0\}$.

Proposition 4.1. *If $X \perp Y$ then $g(X) \perp h(Y)$ for any functions g and h.*

Example 4.4. (Repeat of a previous example) Suppose that we have 10 ActSc students, 9 Stats student and 6 Maths students. We select 5 students without replacement. Let $X = \#$ of ActSc students and $Y = \#$ of Stats students.

Are X and Y independent? (No, they're dependent)

Example 4.5. Let $f_{XY}(x,y) = \frac{3}{2}y(1-x^2)$ for $-1 \le x \le 1$, $0 \le y \le 1$. Are X and Y independent? (Yes, check the marginals)

Example 4.6. Let $f_{XY}(x,y) = \frac{\theta^{x+y}e^{-2\theta}}{x!y!}$. This splits into two independent poisson r.v.s. X and Y.

Example 4.7. Let $f_{XY}(x,y) = \frac{2}{\pi}$ where $0 \le x \le \sqrt{1-y^2}$ and $-1 \le y \le 1$. Calculating the marginals gives us $f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}$ and $f_X(x) = \frac{4}{\pi}\sqrt{1-x^2}$. It is clear that X is not independent of Y.

Remark 4.1. In general, $X_1 \perp X_2 \perp ... \perp X_n$ are independent if and only if

$$f_{X_1X_2...X_n}(x_1,x_2,...,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

Remark 4.2. Given $f_{XYZ}(x,y,z) = g(x,y)h(y,z)$, we remark that $X \perp Z$ if Y is given.

4.1 Joint Expectation and Variance

Proposition 4.2. *If $X_1 \perp X_2 \perp ... \perp X_n$, then*

$$E\prod_{i=1}^{n}[h_i(X_i)] = \prod_{i=1}^{n} E[h_i(X_i)]$$

for any set of equations $\{h_i\}_{i=1}^n$.

Definition 4.2. Define $Cov(X,Y) = E[XY] - E[X]E[Y] = E[XY] - \mu_X\mu_Y$. If $X \perp Y$ then $Cov(X,Y) = 0$. We also say that if $E[XY] = E[X]E[Y]$ then X and Y are *uncorrelated*. However, if for all functions f,g we have that

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$

then $X \perp Y$.

Proposition 4.3. *Suppose that X is uncorrelated to Y and that*

$$Y = \alpha X \implies E[\alpha X^2] = \alpha E[X]E[X] \implies E[X^2] = (E[X])^2 \implies X \text{ is a constant}$$

so X and Y cannot be linearly dependent (still cannot say that they are independent).

Proposition 4.4. *If $X_1 \perp X_2 \perp ... \perp X_n$ then*

$$Var\left[\sum_{i=1}^{n} a_iX_i\right] = \sum_{i=1}^{n} a_i^2 Var[X_i] = \sum_{i=1}^{n} a_i\sigma_{X_i}^2$$

4.2 Correlation Coefficient

Definition 4.3. The *correlation coefficient* ρ for two r.v.s is defined as

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}, -1 \le \rho_{XY} \le 1$$

Example 4.8. Recall the pdf $f_{XY}(x,y) = x+y$ on $0 \le x,y \le 1$ and 0 otherwise. We showed that the marginals were $f_X(x) = x + \frac{1}{2}$, $f_Y(y) = y + \frac{1}{2}$ for $0 \le x,y \le 1$ and 0 otherwise. It can be shown that

$$E[XY] = \frac{1}{3}, E[X] = \frac{7}{12} = E[Y], Var(X) = Var(Y) = \frac{11}{144} \implies \sigma_X = \sigma_Y = \frac{\sqrt{11}}{12}$$

and so $\rho = \frac{\frac{1}{3} - \frac{7}{12}\times\frac{7}{12}}{\frac{\sqrt{11}}{12}\times\frac{\sqrt{11}}{12}} = -\frac{1}{11}$.

Proposition 4.5. $-1 \le \rho_{XY} \le 1$ *for any r.v.s X and Y.*

Proof. Consider

$$E\left[\left(\frac{X-\mu_X}{\sigma_X}+\frac{Y-\mu_Y}{\sigma_Y}\right)^2\right] = \frac{1}{\sigma_X^2}E\left[(X-\mu_X)^2\right]+\frac{1}{\sigma_Y^2}E\left[(Y-\mu_Y)^2\right]+\frac{2}{\sigma_X\sigma_Y}E[(X-\mu_X)(Y-\mu_Y)]$$
$$= \frac{\sigma_X^2}{\sigma_X^2}+\frac{\sigma_Y^2}{\sigma_Y^2}+\frac{2}{\sigma_X\sigma_Y}Cov(X,Y)\geq 0$$

and so $\frac{Cov(X,Y)}{\sigma_X\sigma_Y}\geq -1$. A similar method can be constructed using $E\left[\left(\frac{X-\mu_X}{\sigma_X}-\frac{Y-\mu_Y}{\sigma_Y}\right)^2\right]$ in the above to get $\frac{Cov(X,Y)}{\sigma_X\sigma_Y}\leq 1$. □

5 Conditional Distributions

Definition 5.1. For r.v.s X and Y,

$$f_{X|Y}(x|y)=\frac{f_{XY}(x,y)}{f_Y(y)}, p_{X|Y}(x|y)=\frac{p_{XY}(x,y)}{p_Y(y)} \text{ and } f_Y,p_Y>0$$

Example 5.1. Consider $f_{XY}(x,y)=\frac{2}{\pi}$ on $-1\leq y\leq 1$, $0\leq x\leq\sqrt{1-y^2}$ and 0 otherwise. We computed the marginals to be $f_X(x)=\frac{4}{\pi}\sqrt{1-x^2}$ and $f_Y(x)=\frac{2}{\pi}\sqrt{1-y^2}$. It is easy to show that

$$f_{X|Y}(x|y)=\frac{1}{\sqrt{1-y^2}}, f_{Y|X}(y|x)=\frac{1}{2\sqrt{1-x^2}}$$

Remark 5.1. (Product Rule) We can express the joint in the following way

$$f_{XY}(x,y)=f_{X|Y}(x|y)\cdot f_Y(y)=f_{Y|X}(y|x)\cdot f_X(x)$$

Example 5.2. Suppose that $Y\sim Pois(\mu)$ and $X|Y_{=y}\sim Bin(y,p)$. What is the marginal of X? The joint distribution is

$$p_{XY}(x,y)=\frac{e^{-\mu}\mu^y}{y!}\left[\frac{y!}{x!(y-x)!}p^x(1-p)^{y-x}\right]=\frac{e^{-\mu}(\mu p)^x(\mu(1-p))^{y-x}}{x!(y-x)!}$$

so

$$PHp_X(x) = \sum_{y=x}^{\infty}\frac{e^{-\mu}(\mu p)^x(\mu(1-p))^{y-x}}{x!(y-x)!}$$
$$= \frac{e^{-\mu}}{x!}(\mu p)^x\sum_{y=x}^{\infty}\frac{(\mu(1-p))^{y-x}}{(y-x)!}$$
$$= \frac{e^{-\mu}e^{(\mu(1-p))}}{x!}(\mu p)^x$$
$$= \frac{e^{-\mu p}(\mu p)^x}{x!}$$

and $X\sim Pois(\mu p)$.

Example 5.3. Given $P_{XY}(x,y)=\frac{\theta^{x+y}e^{-2\theta}}{x!y!}$ for $x,y=0,1,2,...$ It can be shown that

$$P_X(x)=\sum_{y=0}^{\infty}P_{XY}(x,y)=\frac{\theta^x e^{-\theta}}{x!}, P_{Y|X}(y|x)=\frac{\theta^y e^{-\theta}}{y!}$$

Example 5.4. Suppose that $Y \sim$ Gamma $(\alpha, \frac{1}{\theta})$ and $X|Y_{=y} \sim$ Wei$(y^{-\frac{1}{p}}, p)$. What is f_X? We know that

$$\text{Gamma}(\alpha,\beta) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)}, \text{Wei}(\theta,\beta) = \frac{\beta}{\theta^\beta} x^{\beta-1} e^{-\left(\frac{x}{\beta}\right)^\theta}, x \geq 0$$

It can be shown that

$$f_Y(y) = \frac{\theta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\theta y}, f_{X|Y}(x|y) = \frac{p}{\left(y^{-\frac{1}{p}}\right)^p} x^{p-1} e^{-\left(\frac{x}{y^{-\frac{1}{p}}}\right)^p} = pyx^{p-1}e^{-yx^p}, x \geq 0, y \geq 0$$

This gives an equation for f_{XY} in the form of

$$f_{XY}(x,y) = \frac{py\theta^\alpha}{\Gamma(\alpha)} x^{p-1} y^{\alpha-1} e^{-\theta y} e^{-yx^P}$$

and integrating gives us

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{XY}(x,y)\,dy \\ &= \frac{p\theta^\alpha x^{p-1}}{\Gamma(\alpha)(\theta+x^p)^{\alpha+1}} \int_0^\infty t^\alpha e^{-t}\,dt, t = y(\theta+x^p) \\ &= \frac{\Gamma(\alpha+1)p\theta^\alpha x^{p-1}}{\Gamma(\alpha)(\theta+x^p)^{\alpha+1}} \end{aligned}$$

5.1 Conditional Expectation

Definition 5.2. $E[g(Y)|x] = \sum_y g(y) f_{XY}(y|x) = E[g(Y)|X=x]$ and if $X \perp Y$ then $E[g(Y)|X=x] = E[g(Y)]$. Variance is defined in a similar way: $Var[Y|X=x] = E[Y^2|X=x] - E^2[Y|X=x]$.

Example 5.5. Let $f_{Y|X}(y|x) = \frac{1}{2\sqrt{1-x^2}}, -\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2}$. We want to compute the variance. First,

$$E[Y|X=x] = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{y}{2\sqrt{1-x^2}}\,dy = 0$$

since the term in the integral is an odd function. Then,

$$E[Y^2|X=x] = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{y^2}{2\sqrt{1-x^2}}\,dy = \frac{1}{\sqrt{1-x^2}} \int_0^{\sqrt{1-x^2}} y^2\,dy = \frac{(1-x^2)^3}{3\sqrt{1-x^2}} = \frac{1}{3}(1-x^2)$$

So $Var[Y|X=x] = \frac{1}{3}(1-x^2)$.

Proposition 5.1. *The double expectation formula states*

$$E[X] = E\left(E[X|Y]\right) \implies Var[X] = E[Var(X|Y)] + Var(E[X|Y)$$

Example 5.6. If $P \sim Unif([0,1])$ and $Y|P_{=p} \sim Bin(10,p)$ then

$$E[E[Y|P]] = E[10P] = 10E[P] = \frac{10}{2} = 5$$

and

$$Var[E[Y|P] = Var[10P] = \frac{100}{12}$$

(Some examples we skip here because they are trivial)

Example 5.7. Given

$$f_{XY} = \begin{cases} 6xy(2-x-y) & 0<x,y<1 \\ 0 & \text{otherwise} \end{cases}$$

we can show that $E[XY] = \frac{1}{3}$, $f_{X|Y} = \frac{6x(2-x-y)}{4-3y}$ with $f_Y = y(4-3y), 0<y<1$

Definition 5.3. The joint MGF of XY is defined as

$$M_{XY}(t) = E\left[e^{tX+tY}\right]$$

and in general,

$$M_{\prod_{i=1}^{k} X_i}(t_1,..,t_k) = E\left[e^{\sum_{l=1}^{k} t_1 X_i}\right]$$

with

$$M_{\prod_{i=1}^{k} X_i}(t_1, t_2 = 0, ..., t_k) = E\left[e^{t_1 X_1}\right] = M_{X_1}(t)$$

Proposition 5.2. *As $y \to \infty$, $F_{XY}(x,y) \to F_X(x)$.*

Example 5.8. Given

$$f_{XY} = \begin{cases} e^{-y} & 0<x<y<\infty \\ 0 & \text{otherwise} \end{cases}$$

we can show that

$$M_{XY}(t_1,t_2) = \frac{1}{(t_1+t_2-1)(t_2-1)}$$

Proposition 5.3. *If $\{X_k\}$ are a set of independent random variables,*

$$M_{\prod X_k}(t_1,...,t_n) = \prod M_{X_k}(t_k)$$

Exercise 5.1. Show that if $X_1,...,X_n$ are iid $N(0,1)$ r.v.s, then $Y = \sqrt{n}\bar{X}_n \sim N(0,1)$. First remark that $Var(Y) = Var(n\left[\frac{\sum X_n}{\sqrt{n}}\right]) = 1$. We can further calculate the MGF of Y as

$$M_Y(t) = \prod E\left[e^{\frac{x_1 t}{\sqrt{n}}}\right] = \prod e^{\frac{t^2}{2n}} = \left(e^{\frac{t^2}{2n}}\right)^n = e^{\frac{t^2}{2}}$$

and so $Y = \sqrt{n}\bar{X}_n \sim N(0,1)$.

Theorem 5.1. *If $Y_1,...,Y_n \sim N(0,1)$ and they are independent, then*

$$\frac{\bar{Y}_n - \mu}{\sigma^2/\sqrt{n}} \sim N(0,1)$$

6 Multivariable Distributions

Here, we examine various distributions that comprise of multiple variables.

6.1 Multinomial Distribution

Definition 6.1. Let X_i be the number of times "i" comes before n total repetitions, and p_i be the probability of getting the item "i". Then

$$P(X_1 = x_1,...,X_k = x_k) = \frac{n!}{x_1!...x_k!} p_1^{x_1} \cdot ... \cdot p_k^{x_k}$$

where $\sum x_i = n, \sum p_i = 1$. We say that $(X_1,...,X_k) \sim Mult(n,p_1,...,p_k)$.

Proposition. *Some properties include:*

1. $M_X(t) = (p_1 e^{t_1} + ... + p_k^{t_k} + p_{k+1})^n$
2. $Cov(X_i, X_j) = -np_i p_j$

6.2 Bivariate Normal Distribution

Definition 6.2. If X_1 and X_2 have the following joint PDF:

$$f_{X_1X_2}(x_1,x_2) = \frac{1}{2\pi|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)\right\}, x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

then $X = (X_1, X_2)^t \sim BivN(\mu, \Sigma)$. Note that the matrix Σ must be positive definite.

Remark 6.1. If $\rho = 0$ then

$$\begin{aligned} f_{X_1X_2} &= \frac{1}{2\pi\sigma_1\sigma_2} exp\left\{-\frac{1}{2\sigma_1^2\sigma_2^2}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}^t \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}\right\} \\ &= \underbrace{\frac{1}{\sqrt{2\pi}\sigma_1} e^{\left(-\frac{x_1-\mu_1}{2\sigma_1^2}\right)}}_{N(\mu_1,\sigma_1^2)} \cdot \underbrace{\frac{1}{\sqrt{2\pi}\sigma_2} e^{\left(-\frac{x_2-\mu_2}{2\sigma_2^2}\right)}}_{N(\mu_2,\sigma_2^2)} \end{aligned}$$

So X_1 and X_2 are independent. This is special to only the bivariate normal r.v.

Note 2. In general, if $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ then if $\rho_{X_1X_2} = 0$ it is not always true that $X_1 \perp X_2$. This is only the case if X_1 and X_2, collectively, are bivariate normal.

Summary 6. Here are some values that may be useful in the computation of f_{X_1,X_2}:

$$|\Sigma| = \sigma_1^2\sigma_2^2(1-\rho^2), |\Sigma|^{1/2} = \sigma_1\sigma_2\sqrt{1-\rho^2}$$

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

and so

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left\{-\frac{1}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\left[(x_1-\mu_1)^2\sigma_2^2 - 2(x_1-\mu_1)(x_2-\mu_2)\rho\sigma_1\sigma_2 + (x_2-\mu_2)^2\sigma_1^2\right]\right\} \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\} \end{aligned}$$

Fact 6.1. *The moment generating function is*

$$M_X(t_1,t_2) = E\left[e^{t^tX}\right] = E\left[e^{t_1X_1+t_2X_2}\right] = ... = e^{\mu^t t + \frac{1}{2}t^t\Sigma t}$$

where $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ *and*

$$M_{X_1} = M_X(t_1, 0), M_{X_2} = M_X(0, t_2)$$

Proposition 6.1. *If* $C = \begin{pmatrix} C_1 & C_2 \end{pmatrix}^t$ *then* $C^tX \sim N(C^t\mu, C^t\Sigma C)$ *and* $Y = AX + b \implies Y \sim N(A\mu + b, A\Sigma A^t)$.

Remark 6.2. For a condition distribution $X_2|X_1 = x_1$ with X_2, X_1 being jointly bivariate, we have

$$X_2|X_1 = x_1 \sim N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1), \sigma_2^2(1-\rho^2))$$

This can be done by putting the joint over of the marginal of X_1. For the sake of sanity, I will not be bashing through the computation of this.

Fact 6.2. $E[X_1X_2] = E[E[X_1X_2|X_2]]$

Example 6.1. Suppose that X_1X_2 are $BIV(\mu, \Sigma)$. Then

$$E[X_1X_2|X_2 = x_2] = x_2E[X_1|X_2 = x_2] = x_2(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2))$$

and so $E[X_1X_2|X_2] = X_2(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2))$. Thus,

$$E[X_1X_2] = E[E[X_1X_2|X_2]] = E[X_2](\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(E[X_2] - \mu_2)) = \mu_1\mu_2 + \rho\sigma_1\sigma_2$$

and so we can represent the covariance of X_1 and X_2 as

$$Cov(X_1, X_2) = \mu_1\mu_2 + \rho\sigma_1\sigma_2 - \mu_1\mu_2 = \rho\sigma_1\sigma_2$$

7 Functions of Random Variables

Example 7.1. Suppose that $X = Z^2$ and $f_Z(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$. Remark that

$$F_X(x) = P(X \leq x) = P(Z^2 \leq x) = P(-\sqrt{x} \leq Z \leq \sqrt{x}) = F_Z(\sqrt{x}) - F_Z(-\sqrt{x})$$

So taking derivatives, we have

$$f_X(x) = \frac{1}{x^{1/2}}f_Z(x) = \frac{1}{\sqrt{2\pi}x^{1/2}}e^{-x/2}$$

and note that $X \sim$ Gam$(2, \frac{1}{2})$.

Fact 7.1. *If $Z_1, ..., Z_n$ are independent $N(0, 1)$ then*

$$X = Z_1^2 + ... + Z_n^2 = \chi_n^2$$

and $E[X] = n$.

Example 7.2. Suppose that

$$f_{XY}(3y) = \begin{cases} 3y & 0 \leq x \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the pdf of $T = XY$. Now since $P(T \leq t) = P(XY < t) = 1 - P(XY > t)$ then we calculate $P(XY > t)$ as

$$P(XY > t) = \int_{\sqrt{t}}^{1}\int_{\frac{t}{y}}^{y} 3y\,dx\,dy = 1 + 2t\sqrt{t} - 3t$$

What is the pdf of T? By direct computation, this is

$$f_T = 3 - 3t^{\frac{1}{2}}, 0 \leq t \leq 1$$

What is the pdf of $S = \frac{Y}{X}$. Well, the cdf is

$$F_S(s) = \int_0^1\int_{\frac{y}{s}}^{y} 3y\,dx\,dy = 1 - \frac{1}{s}$$

and so the pdf is

$$\frac{\partial}{\partial s}F_S(s) = f_S(s) = \frac{1}{s^2}$$

Example 7.3. Suppose that $X_1, ..., X_n$ are iid with pdf f_X and cdf F_X. Let $Y = \max(X_1, .., X_n)$ and $T = \min(X_1, .., X_n)$. So

$$F_Y(y) = P(Y \leq y) = P(X_1 \leq y, ..., X_n \leq y) = \prod_{i=1}^{n} F_{X_i}(y) = F_X^n(y) \implies f_Y(y) = nf_X(y)F_x^{n-1}(y)$$

and

$$F_T(t) = P(T < t) = 1 - P(T \geq t) = 1 - \prod_{i=1}^{n}(1 - F_{X_i}(t)) = 1 - (1 - F_X(t))^n \implies f_T(t) = n(1 - F_X(t))^{n-1} f_X(t)$$

Example 7.4. If each X_i was $exp(\lambda_i)$ then $F_{X_i}(x) = 1 - e^{-\lambda_i x}$ and so

$$1 - F_T(t) = e^{-(\sum_{i=1}^{n}\lambda_i)t}, t \geq 0 \implies F_T(t) = 1 - e^{-(\sum_{i=1}^{n}\lambda_i)t}, t \geq 0$$

and $T \sim exp(\sum_{i=1}^{n}\lambda_i)$.

Example 7.5. Suppose that Z_1 and Z_2 are i.i.d. r.v.s that are $N(0,1)$. What is the distribution of $X = \frac{(Z_1-Z_2)^2}{2}$? Well, note that

$$\frac{Z_1 - Z_2}{\sqrt{2}} \sim N(0,1)$$

so

$$\left(\frac{Z_1 - Z_2}{\sqrt{2}}\right)^2 \sim \chi_1^2$$

7.1 1-to-1 Bivariate Transformations

If we are given an (X, Y) bivariate vector (2 r.v.s) and $f_{X,Y}(x, y)$ is known, then let

$$A = \{(x, y), f_{XY} > 0\}, B = \{(u, v), u = h_1(x, y), v = h_2(x, y)\}$$

If $U = h_1(X, Y)$, $V = h_2(X, Y)$ and $X = w_1(U, V), Y = w_2(U, V)$ then

$$g_{UV}(u, v) = f_{XY}(x, y) \cdot \left|\frac{\partial(x, y)}{\partial(u, v)}\right| = f_{XY}(w_1(u, v), w_2(u, v)) \cdot \left|\frac{\partial(x, y)}{\partial(u, v)}\right|$$

Example 7.6. Suppose that

$$f_{XY}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2+y^2}{2}}$$

and

$$U = X + Y, V = X - Y \implies X = \frac{U + V}{2}, Y = \frac{U - V}{2}$$

then

$$g_{UV}(u, v) = \frac{1}{\sqrt{2\pi \times 2}} e^{-\frac{u^2+v^2}{4}} = \frac{1}{\sqrt{2\pi \times 2}} e^{-\frac{u^2}{2\times 2}} \frac{1}{\sqrt{2\pi \times 2}} e^{-\frac{v^2}{2\times 2}}$$

Example 7.7. Suppose that $X \sim Unif([0,1))$ and $Y \sim Unif([0,1))$. Using the *Box-Muller transformation*, if

$$U = \sqrt{-2\ln X}\,(\cos 2\pi Y)\,, V = \sqrt{-2\ln X}\,(\sin 2\pi Y)$$

it can be shown that U, V are independent $N(0,1)$. Also the Jacobian is $J = -\frac{x}{2\pi}$.

Now note that $U^2 + V^2 = -2\ln x \implies X = \exp\left(-\frac{U^2+V^2}{2}\right)$ so $|J| = \frac{1}{2\pi} e^{-\frac{U^2+V^2}{2}}$.

Example 7.8. Suppose that we have

$$f_{XY}(x, y) = \begin{cases} e^{-y} & 0 < x < y < \infty \\ 0 & \text{o/w} \end{cases}$$

If $U = X + Y$ and $V = X$. Then $X = V$, $Y = U - V$ and the support is $0 < 2v < u < \infty$. Our Jacobian is

$$|J| = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = 1$$

and so

$$g_{UV}(u,v) = \begin{cases} e^{-(u-v)} & 0 < 2v < u < \infty \\ 0 & \text{o/w} \end{cases} \implies g_U(u) = \int_0^{u/2} e^{-u}e^{v}dv = e^{-u/2} - e^{-u}$$

$$\implies g_V(v) = \int_0^{2v} e^{-u}e^{v}du = e^{v}(-e^{-2v}-1) = -e^{-v} - e^{v}$$

Example 7.9. Suppose that we have

$$f_{XY}(x,y) = \begin{cases} e^{-x-y} & 0 < x, y < \infty \\ 0 & \text{o/w} \end{cases}$$

and $U = X+Y$, $V = X$ then $X = V$ and $Y = U - V$ with $|J| = 1$. The support is $0 < v < u < \infty$ and $g_{UV}(u,v) = e^{-v-(u-v)} = e^{-u}$.

Definition 7.1. If $Z \sim N(0,1)$, $X \sim \chi^2_n$, and $Y \sim \chi^2_m$ then $Z/\sqrt{\frac{X}{n}} \sim t_n$ and $\frac{X/n}{Y/m} \sim F_{n,m}$.

Remark 7.1. If $W \sim F_{n,m}$ then $V = \frac{1}{W} \sim F_{m,n}$.

Example 7.10. To compute the pdf of t_n let $U = X$ and $V = Z/\sqrt{\frac{X}{n}}$. Then $X = U$ and $Z = \frac{V}{\sqrt{n}}\sqrt{U}$. We can use the Jacobian method above to compute the pdf.

7.2 Moment-Generating Function Method

Fact 7.2. *If $X_1, X_2, ..., X_n$ are independent and X_i has MGF $M_{X_i}(t)$ then if $Y = \sum_{i=1}^n X_i$ we have*

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t)$$

and if the $X_i's$ are i.i.d. then

$$M_Y(t) = M^n_{X_1}(t)$$

Example 7.11. Suppose that $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$ where $M_X(t) = e^{\mu t}e^{\sigma^2 t^2/2}$. Then

$$M_Y(t) = E[e^{tY}] = e^{bt}E[e^{(at)X}] = e^{bt}e^{a\mu t + \frac{\sigma^2 a^2 t^2}{2}} = e^{(a\mu+b)t}e^{\frac{a^2\sigma^2 t^2}{2}} \implies Y \sim N(a\mu+b, a^2\sigma^2)$$

Now if $X_i \sim N(\mu_i, \sigma_i^2)$ and $Y = \sum_{i=1}^n a_i X_i$ then

$$M_Y(t) = E[e^{t\sum_{i=1}^n a_i X_i}] = \prod_{i=1}^n E[e^{ta_iX_i}] = \prod_{i=1}^n e^{a_i\mu_i t}e^{\frac{a_i^2\sigma_i^2 t^2}{2}} = e^{t\sum_{i=1}^n a_i\mu_i}e^{\frac{t^2}{2}\sum_{i=1}^n a_i^2\sigma_i^2} \implies Y \sim N\left(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\mu_i^2\right)$$

Corollary 7.1. *Suppose that we have $X_i \sim N(\mu, \sigma^2)$ where $X_1, ..., X_n$ are i.i.d. then*

$$\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2), \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \sim N(\mu, \frac{\sigma^2}{n})$$

Fact 7.3. *We have*

$$Y = \sum_{i=1}^n \chi^2_{m_i} = \chi^2_{\sum_{i=1}^n m_i}$$

Example 7.12. We know that

$$\underbrace{\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma^2}\right)^2}_{\chi^2_{(n)}} = \underbrace{\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma^2}\right)^2}_{\chi^2_{(n-1)} \text{ from other 2}} + \underbrace{n\left(\frac{X_i - \bar{X}}{\sigma^2}\right)^2}_{\chi^2_{(1)}}$$

Proof is an exercise.

Corollary 7.2. *$\bar{X}$ and $S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$ are independent. (Corchan's Theorem)*

Fact 7.4. *We have that if $X_i \sim N(\mu, \sigma^2)$ and $\bar{X}$ and S^2 are defined as above, then*

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{(n-1)}$$

Fact 7.5. *If $X_i \sim N(\mu_1, \sigma_1^2)$ and $Y_j \sim N(\mu_2, \sigma_2^2)$ are i.i.d. for $i = 1, ..., n$ and $j = 1, ..., m$ then*

$$\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F$$

8 Convergence of Random Variables

Convergence can take place (from strongest to weakest):

- Everywhere
- Almost surely in $L^1, L^2, ...$ (See PMATH 450)
- In distribution
- In probability

We will examine the last two definitions of convergence.

Definition 8.1. The sequence $X_1, X_2, ..., X_n$ *converges in probability* to X if for any $\epsilon > 0$ we have

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0 \iff \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$$

We denote this by $X_n \xrightarrow{p} X$.

Definition 8.2. We say that $\{X_n : \Omega \mapsto A_n\}$ *converges in distribution* to $X : \Omega \mapsto B$ if for any $\epsilon > 0$ we have

$$\lim_{n\to\infty} \left|P(X^{-1}(k)) - P(X_n^{-1}(k))\right| = \lim_{n\to\infty} |P(X) - P(X_n)| < \epsilon, k \in A_n \cap B$$

An example would be the central limit theorem. Alternatively, this is equivalent to

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all parts where $F_X(x)$ is continuous. We then write $X_n \xrightarrow{d} X$.

Proposition 8.1. *If $X_n \xrightarrow{p} X$ then $X_n \xrightarrow{d} X$.*

Example 8.1. Suppose that $\{X_k\}_{k=1}^n$ are i.i.d. $Unif([0,1])$. Let $X_{(n)} = \max_k X_k$ and $X_{(1)} = \min_k X_k$.

(1) What is the limiting distribution of $nX_{(1)}$?

First remark that the support of $nX_{(1)}$ is $(0, n)$. Then note that for $0 < x < n$ we have

$$P(nX_{(1)} \leq x) = 1 - P\left(X_{(1)} > \frac{x}{n}\right) = 1 - \prod_{k=1}^{n} P\left(X_k > \frac{x}{n}\right) = 1 - \left(1 - \frac{x}{n}\right)^n$$

So

$$F_n(x) = P(nX_{(1)} \le x) = \begin{cases} 0 & x \le 0 \\ 1 - \left(1 - \frac{x}{n}\right)^n & 0 < x < n \\ 1 & x \ge n \end{cases} \implies \lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-x} & x > 0 \end{cases}$$

(2) What is the limiting distribution of $n(1 - X_{(n)})$?

Similar to above, the support of $n(1 - X_{(n)})$ is $(0, n)$ and

$$P(n(1-X_{(n)}) \le x) = 1 - P(n(1-X_{(n)}) > x) = 1 - P\left(X_{(n)} \le 1 - \frac{x}{n}\right) = 1 - \prod_{k=1}^{n} P\left(X_k \le 1 - \frac{x}{n}\right) = 1 - \left(1 - \frac{x}{n}\right)^n$$

and in the limit, we have the same distribution in (1). That is

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-x} & x > 0 \end{cases}$$

(3) What is the limiting distribution of $X_{(1)}$?

This can be shown to have limiting cdf of

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x < 0 \\ 1 & x \ge 0 \end{cases} \implies X_{(1)} \xrightarrow{d} 0$$

(4) Similarly, what is the limiting distribution of $X_{(n)}$?

This can be shown to have limiting cdf of

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x < 1 \\ 1 & x \ge 1 \end{cases} \implies X_{(1)} \xrightarrow{d} 1$$

Definition 8.3. Given a sequence of r.v.s. $\{X_n\}_{n=1}^{\infty}$, with corresponding cdfs $\{F_n(x)\}_{n=1}^{\infty}$ if

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & x < b \\ 1 & x \ge b \end{cases}$$

then $X_n \xrightarrow{d} b$.

Theorem 8.1. *If $X_n \xrightarrow{d} b$ then $X_n \xrightarrow{p} b$.*

Proof. By direct evaluation,

$$\begin{aligned} P(|X_n - b| > \epsilon) &= P(X_n < b - \epsilon) + P(X_n > b + \epsilon) \\ &\le P(X_n \le b - \epsilon) + P(X_n > b + \epsilon) \\ &= F_n(b - \epsilon) + 1 - F_n(b + \epsilon) \end{aligned}$$

so taking limits gives us

$$\begin{aligned} \lim_{n\to\infty} P(|X_n - b| > \epsilon) &= \lim_{n\to\infty} F_n(b - \epsilon) + 1 - \lim_{n\to\infty} F_n(b + \epsilon) \\ &= 0 + 1 - 1 \\ &= 0 \end{aligned}$$

□

Example 8.2. Given $\{X_i\}_{i=1}^{\infty}$ i.i.d. r.v.s, with

$$f_{X_i}(x) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & \text{o/w} \end{cases}$$

Let $Y_n = \min_i X_i$ and show that $Y_n \xrightarrow{p} \theta$.

It is easier to show that $Y_n \xrightarrow{d} \theta$. Remark that the support of Y_n is (θ, ∞). We then have

$$P(Y_n \leq x) = \begin{cases} 0 & x < \theta \\ 1 - e^{-n(x-\theta)} & x \geq \theta \end{cases} \implies \lim_{n\to\infty} P(Y_n \leq x) = \begin{cases} 0 & x < \theta \\ 1 & x \geq \theta \end{cases} \implies Y_n \xrightarrow{d} \theta$$

as required.

Fact 8.1. *(Markov's Inequality) For any $k \in \mathbb{N}$,*

$$P(|X| > C) \leq \frac{E[|X|^k]}{C^k} \implies P(|X| > C) \leq \frac{E[|X|^2]}{C^2} = \frac{Var(X) + (E[X])^2}{C^2}, k = 2$$

Proposition 8.2. *A property of the arithmetic mean of random variables is $\bar{X} \xrightarrow{p} \mu$.*

Proof. We have

$$0 \leq P(|\bar{X} - \mu| > \epsilon) \leq \frac{1}{\epsilon^2} E[(\bar{X} - \mu)^2] = \frac{1}{\epsilon^2} Var(\bar{X}) \leq \frac{\sigma^2}{n\epsilon^2} \to 0$$

□

Remark 8.1. If $X_n = 1 - X$ then $X_n \xrightarrow{d} X$ but $X_n \not\xrightarrow{p} X$.

Theorem 8.2. *(Central Limit Theorem) $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{d} N(0,1)$ where $\{X_n\}$ are i.i.d. r.v.s. with $X_n \sim (\mu, \sigma^2)$*

Proof. (No. 1) Observe that for the cdf of any r.v. X we have $M_X(t) = e^{t^2/2}$, $f(0) = \ln M_X(0) = 0$,

$$f'(0) = \frac{M_X'(0)}{M_X(0)} = 0, f''(0) = \frac{M_X''(0) M_X(0) - (M_X(0))^2}{M_X^2(0)} = 1$$

Now if $A = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$ and $Y_i = \frac{X_i - \mu}{\sigma}$ then

$$E[e^{tA}] = E\left[e^{\frac{t}{\sqrt{n}} \sum_{i=1}^n Y_i}\right] = \prod_{i=1}^n e^{\frac{t}{\sqrt{n}} Y_i} = M_Y^n\left(\frac{t}{\sqrt{n}}\right)$$

since

$$\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right) = \frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^n X_i - n\mu\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i - \mu}{\sigma}$$

Now from above,

$$f_Y\left(\frac{t}{\sqrt{n}}\right) = \ln M_Y\left(\frac{t}{\sqrt{n}}\right) = \frac{t^2}{2n} + O(t^3) \implies M_Y\left(\frac{t}{\sqrt{n}}\right) = e^{t^2/2n} \implies M_Y^n\left(\frac{t}{\sqrt{n}}\right) = e^{t/2n}$$

and hence $\lim_{n\to\infty} M_Y\left(\frac{t}{\sqrt{n}}\right) = e^{t^2/2} \xrightarrow{d} N(0,1)$. □

Proof. (No. 1) Alternatively, using notation from the previous proof,

$$M_A(t) = M_{\sum_{i=1}^n Y_i/\sqrt{n}}(t) = M_Y^n\left(\frac{t}{\sqrt{n}}\right)$$

Using a 1st order Taylor series,

$$\lim_{n\to\infty} M_Y^n\left(\frac{t}{\sqrt{n}}\right) = \lim_{n\to\infty}\left(1+\frac{t^2}{2n}\right)^n = e^{t^2/2}$$

□

Corollary 8.1. *(1) If* $\{X_i\}$ *are i.i.d.* $Pois(\mu)$ *and* $Y_n = \sum_{i=1}^n X_i$ *then*

$$\frac{Y_n - n\mu}{\sqrt{n\mu}} \xrightarrow{d} N(0,1)$$

(2) If $\{X_i\}$ *are i.i.d.* χ_1^2 *(mean of* χ_k^2 *is* k *and variance is* $2k$*) and* $Y_n = \sum_{i=1}^n X_i$ *then*

$$\frac{Y_n - n}{\sqrt{2n}} \xrightarrow{d} N(0,1)$$

8.1 Useful Limit Theorems

1. If $X_n \overset{p\ (D)}{\rightarrow} a$ then $g(X_n) \overset{p\ (D)}{\rightarrow} g(a)$. That is g is continuous at "a".
2. (Slutsky's Theorem) Suppose that $X_n \xrightarrow{d} X$ and $Y \xrightarrow{p} b$. Then,
   (a) $X_n + Y_n \xrightarrow{d} X + b$
   (b) $X_n \cdot Y_n \xrightarrow{d} b \cdot X$
   (c) $X_n/Y_n \xrightarrow{d} X/b$, $b \neq 0$

Example 8.3. Suppose $X_1, X_2, ..., X_n$ are i.i.d $X_i \sim Unif[0,1)$. We showed that $X_{(n)} \xrightarrow{p} 1 \implies e^{X_{(n)}} \xrightarrow{p} e$ and $n(1-X_{(n)}) \xrightarrow{d} Z \sim exp(1)$.

Remark 8.2. $F_X(X) \sim Unif([0,1])$

Example 8.4. If $\{X_i\}_{i=1}^n$ are $Pois(\mu)$ then

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\bar{X}_n}} = \underbrace{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\bar{X}_n}}}_{\xrightarrow{d} N(0,1)} \underbrace{\frac{\sqrt{\bar{X}_n}}{\sqrt{\mu}}}_{\xrightarrow{p} 1} \sim N(0,1)$$

and similarly

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0,\mu)$$

8.2 Delta Method

Proposition 8.3. *Suppose that for* $X_1, X_2, ..., X_n$ *we have*

$$\sqrt{n}(X_n - \theta) \xrightarrow{d} N(0,\sigma)$$

If $g(x)$ *is differentiable at* θ *and* $g'(\theta) \neq 0$ *then*

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} N(0, g'(\theta)^2\sigma^2)$$

9 Point Estimation

Suppose that we observe $X_1, ..., X_n$ i.i.d. from $f(x, \theta)$ and θ is unknown. The goal is to estimate θ.

Definition 9.1. The $t-$statistic is a function of data that doesn't depend on θ or μ (or any unknown parameter). We denote it by $T(X) = T(X_1, ..., X_n)$ as a random variable and $t = t(x_1, ..., x_n)$ as its value.

The following are different methods for point estimation.

1. Method of Moments
2. Maximum Likelihood
3. Bayes Estimation

9.1 Method of Moments

Here, we want to set the sample/observed k^{th} moment equal to the theoretical moment. That is we want

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \longleftrightarrow E[X^l]$$

for $l \in \{1, 2, ..., l\}$.

Example 9.1. If $X_1, ..., X_n$ are i.i.d. $Pois(\mu)$ then $E[X_i] = \mu$ and

$$\hat{\mu}_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Example 9.2. If $X_1, ..., X_n$ are i.i.d. and

$$f_X = \begin{cases} \frac{1}{\theta}e^{-x/\theta} & x \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

Then $E[X_i] = \theta$ and

$$\hat{\theta}_{MM} = \frac{1}{N}\sum_{i=1}^{n} X_i$$

Example 9.3. If $X_1, ..., X_n$ are i.i.d. $N(\mu, \sigma^2)$ then

$$\hat{\mu}_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i, \hat{\sigma}^2_{MM} + \hat{\mu}^2_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Example 9.4. If $X_1, ..., X_n$ are i.i.d. and

$$f_X = \begin{cases} \theta e^{\theta - 1} & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}, \theta > 0$$

Then we can show

$$E[X] = \frac{\theta}{\theta + 1} \implies \theta = \frac{E[X]}{1 - E[X]} \implies \hat{\theta}_{MM} = \frac{\bar{X}}{1 - \bar{X}}$$

Example 9.5. Suppose that $X \sim \text{Gam}(\alpha, \beta)$ then $E[X] = \alpha\beta$ and $Var(X) = \alpha\beta^2$. So

$$\hat{\alpha\beta}_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i, \hat{\alpha\beta^2}_{MM} + \left(\hat{\alpha\beta}_{MM}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Example 9.6. If $X_1, ..., X_n$ are i.i.d. $Unif([0,\theta])$ then

$$\hat{\theta}_{MM} = \frac{2}{n}\sum_{i=1}^{n} X_i, E\left[\hat{\theta}_{MM}\right] = \theta$$

Remark that $\hat{\theta}_{MLE} = \max(X_1, ..., X_n)$.

9.2 Maximum Likelihood Estimation

Definition 9.2. Suppose that $X_1, ..., X_n$ are i.i.d. from $f(x,\theta)$. We call

$$L(\theta, X) = \prod_{i=1}^{n} f(x_i, \theta)$$

the *likelihood* of θ and $l = \ln(L)$ the *log-likelihood* function. The MLE estimate is

$$\hat{\theta}_{ML} = \hat{\theta}_{MLE} = \text{argmax } L(\theta) = \text{argmax } l(\theta)$$

Example 9.7. If $X_1, ..., X_n$ are i.i.d. and

$$f_X = \begin{cases} \frac{1}{\theta}e^{-x/\theta} & x \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

Then it can be shown that

$$l(\theta) = -\ln\theta^n + \ln e^{-(x_1+...+x_n)/\theta} = -n\ln\theta - \frac{x_1 + ... + x_n}{\theta}$$

If we set $\frac{\partial l(\theta)}{\partial\theta} = 0$ then

$$\hat{\theta}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Example 9.8. Suppose that $X_1, ..., X_n \sim Pois(\theta)$. Then

$$L(\theta) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \implies l(\theta) = -n\theta + \ln\theta\sum_{i=1}^{n} x_i - \ln\left(\prod_{i=1}^{n} x_i!\right)$$

and so

$$\frac{\partial l}{\partial\theta} = -n + \frac{\sum_{i=1}^{n} x_i}{\theta} = 0 \implies \hat{\theta}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Example 9.9. Recall that if $f(x,\theta) = \theta x^{\theta-1}, 0 < x < 1, \theta > 1$ then

$$\hat{\theta}_{MM} = \frac{\bar{X}}{1-\bar{X}}$$

It can be shown that

$$L(\theta) = \theta^n\left(\prod_{i=1}^{n} x_i\right)^{\theta-1} \implies l(\theta) = n\ln\theta + (\theta-1)\sum_{i=1}^{n}\ln x_i$$

and so

$$\frac{\partial l}{\partial\theta} = \frac{n}{\theta} + \sum_{i=1}^{n}\ln x_i = 0 \implies \hat{\theta}_{ML} = -\frac{n}{\sum_{i=1}^{n}\ln x_i}$$

Example 9.10. Suppose that $X_1, ..., X_n \sim Exp(1/\theta)$ and i.i.d.. Then

$$L(\theta) = \frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i} \implies l(\theta) = -\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

and hence

$$\frac{\partial l}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i = 0 \implies \hat{\theta}_{ML} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

Example 9.11. Suppose that $X_1, ..., X_n \sim Ber(p)$ and i.i.d.. Then

$$L(p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \implies l(p) = \ln p \sum_{i=1}^{n} x_i + \ln(1-p)\left(n - \sum_{i=1}^{n} xi\right)$$

and so

$$\frac{\partial l}{\partial p} = \frac{\sum_{i=1}^{n}}{p} - \left(n - \sum_{i=1}^{n}\right)\frac{1}{1-p} = 0 \implies \hat{p}_{ML} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Note that

$$\frac{\partial^2 l}{\partial p^2} = -\left[\frac{\sum_{i=1}^{n} x_i}{p^2} + \left(n + \sum_{i=1}^{n} x_i\right)\frac{1}{(1-p)^2}\right] < 0$$

Example 9.12. Suppose that $X_1, ..., X_n$ are i.i.d. $N(\mu, \sigma^2)$. It can be shown that

$$L(\theta) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} \implies l(\theta) = -\frac{n}{2}\ln 2\pi - n\ln\sigma - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

and hence

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0 \implies \hat{\mu}_{ML} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} - +\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \implies \hat{\sigma}_{ML} = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu}_{ML})^2}{n}$$

Example 9.13. Suppose that $X_1, ..., X_n$ are i.i.d. $Unif([0, \theta])$. Then it can be shown that with $f_X = \begin{cases} 1/\theta & 0 \le x \le \theta \\ 0 & \text{o/w} \end{cases}$ we have

$$L(\theta) = \begin{cases} 1/\theta^n & x_1, x_2, .., x_n \in [0, \theta] \\ 0 & \text{o/w} \end{cases} = (1/\theta^n) I_{\max(x_1,...,x_n) \le \theta}$$

and hence $\hat{\theta}_{ML} = \max(x_1, ..., x_n)$.

Proposition 9.1. *$E[(X-a)^2] \ge Var[X]$ and equality holds when $a = E[X]$.*

Proof. We have

$$E[(X - E[X] + E[X] - a)^2] = Var[X] + (E[X]-a)^2 + \underbrace{E\left[(X - E[X])(E[X] - a)\right]}_{=0} \ge Var[X]$$

□

9.3 Notable Functions and Matrices

Definition 9.3. The *score function* is

$$S(\theta) = \frac{\partial}{\partial \theta}\ln f(x, \theta)$$

The *information function* is

$$I(\theta) = \frac{\partial}{\partial \theta}S(\theta) = \frac{\partial^2}{\partial \theta^2}\ln f(x, \theta)$$

The *Fisher information matrix* is

$$J(\theta) = -E\left[I(\theta)\right]$$

Summary 7. Some properties include:

1. $S(\hat{\theta}_{ML}) = 0$
2. $E\left[\frac{\partial}{\partial\theta}\ln f(x,\theta)\right] = 0$
   (a) This follows from the fact that

$$\begin{aligned} E\left[\frac{\partial}{\partial\theta}\ln f(x,\theta)\right] &= \int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\ln f(x,\theta)\cdot f(x,\theta)dx \\ &= \int_{-\infty}^{\infty}\frac{\frac{\partial}{\partial\theta}f(x,\theta)}{f(x,\theta)}\cdot f(x,\theta)dx \\ &= \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}f(x,\theta)dx \\ &= \frac{\partial}{\partial\theta}(1) = 0 \end{aligned}$$

3. $E\left[\frac{\partial^2}{\partial\theta^2}\ln f(x,\theta)\right] = E\left[\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2\right]$
   (a) To see this, we take the partial with respect to θ of $\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\ln f(x,\theta)\cdot f(x,\theta)dx = 0$ to get

$$\int_{-\infty}^{\infty}\frac{\partial^2}{\partial\theta^2}\ln f(x,\theta)\cdot f(x,\theta)dx + \int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\ln f(x,\theta)\cdot\frac{\partial}{\partial\theta}f(x,\theta)dx = 0$$

$$\implies E\left[\frac{\partial^2}{\partial\theta^2}\ln f(x,\theta)\right] - \underbrace{\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\ln f(x,\theta)\cdot\frac{\partial}{\partial\theta}\ln f(x,\theta)\cdot f(x,\theta)dx}_{=E\left[\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2\right]} = 0$$

$$\implies E\left[\frac{\partial^2}{\partial\theta^2}\ln f(x,\theta)\right] = E\left[\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2\right]$$

4. If $X_1, ..., X_n$ are i.i.d. then $E\left[\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2\right] = E\left[\left(\frac{\partial}{\partial\theta}\ln f(x_1,\theta)\right)^2\right]$
   (a) This follows from the definition of $J(\theta)$ with $J(\theta) = nJ_1(\theta)$.

Proposition 9.2. *(Cramer-Rao Lower Bound) Suppose that $T(X_1, ..., X_n)$ is an estimator for θ. Remark that if T is unbiased if $E[T(X)] = \theta$. If $E[T(X)] \neq \theta$ then $E[T(X)]$ is biased. Also, if $X_1, ..., X_n$ are samples from $f(x,\theta)$ then*

$$Var(T) \geq \frac{\left(\frac{\partial}{\partial\theta}E[T]\right)^2}{E\left[\left(\frac{\partial}{\partial\theta}\ln f(x,\theta\right)^2\right]} \geq \frac{1}{E\left[\left(\frac{\partial}{\partial\theta}\ln f(x,\theta\right)^2\right]}$$

Proof. First remark that $Cov(X,Y) \leq Var(X)Var(Y)$. Set $X = T(X)$ and $Y = \frac{\partial}{\partial\theta}\ln f(x,\theta)$. Then

$$\begin{aligned} Cov\left(T(X), \frac{\partial}{\partial\theta}\ln f(x,\theta)\right) &= E\left[T(X)\cdot\frac{\partial}{\partial\theta}\ln f(x,\theta)\right] - \underbrace{E\left[\frac{\partial}{\partial\theta}\ln f(x,\theta)\right]}_{=0}E[T(X)] \\ &= \int_{-\infty}^{\infty}T(X)\frac{\partial}{\partial\theta}f(x,\theta)dx \\ &= \frac{\partial}{\partial\theta}E[T(X)] \end{aligned}$$

Since $Var\left[\frac{\partial}{\partial\theta}\ln f(x,\theta)\right] = E\left[\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2\right]$ because $E\left[\frac{\partial}{\partial\theta}\ln f(x,\theta)\right]^2 = 0$ then

$$Var[T(X)] \geq \frac{\frac{\partial}{\partial\theta}E[T(X)]}{E\left[\left(\frac{\partial}{\partial\theta}\ln f(x,\theta)\right)^2\right]} = \frac{1}{nJ_1(\theta)}$$

□

Example 9.14. Suppose that $X \sim Pois(\mu)$. Then

$$\frac{\partial}{\partial\mu}\ln f(x,\mu) = \frac{\partial}{\partial\mu}\left(-\mu + x\ln\mu - \ln x!\right) = -1 + \frac{x}{\mu}$$

and

$$\frac{\partial^2}{\partial\mu^2}\ln f(x,\mu) = -\frac{x}{\mu^2} \implies I(\mu) = \frac{x}{\mu^2} \implies J(\mu) = \frac{E[X]}{\mu^2} = \frac{1}{\mu}$$

with the C-R (Cramer-Rao) bound as

$$Var(T) \geq \frac{1}{n\frac{1}{\mu}} = \frac{\mu}{n}$$

Previously, we showed that

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i \implies Var[\mu_{ML}] = \frac{1}{n^2}n \cdot \mu = \frac{\mu}{n}$$

and so the ML estimator is efficient.

Remark 9.1. 1) $\hat{\theta}_{ML} \xrightarrow{p} \theta$ (asymptotically)

2) $\sqrt{n}(\hat{\theta}_{ML} - \theta) \xrightarrow{d} N\left(0, \frac{1}{J(\theta)}\right)$ (asymptotically normal)

This will also imply that $\hat{\theta}_{ML} - \theta \xrightarrow{d} N\left(0, \frac{1}{nJ_1(\theta)}\right) = N\left(0, \frac{1}{J(\theta)}\right)$ and $\hat{\theta}_{ML} \to N\left(\theta, \frac{1}{J(\theta)}\right)$.

9.4 Convex Functions

Definition 9.4. We say that a function f is convex if $\forall x_1, x_2 \in (a,b)$ and $\forall\lambda \in [0,1]$ we have

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \iff f'' > 0$$

Remark 9.2. If $-f$ is convex then f is concave.

Proposition 9.3. *(Jensen's inequality) If X is a r.v. and f is convex then*

$$E[f(X)] \geq f\left(E[X]\right)$$

Proof. Suppose that the inequality is true for $k-1 \in \mathbb{N}$. Then

$$\begin{aligned}\sum_{i=1}^{k} p_i f(x_i) &= p_k f(x_k) + (1+p_k)\sum_{i=1}^{k-1}\frac{p_i}{1+p_k}f(x_i) \\ &= p_k f(x_k) + (1-p_k)\sum_{i=1}^{k-1} q_i f(x_i)\end{aligned}$$

and using induction on the latter term we get

$$\begin{aligned}\sum_{i=1}^{k} p_i f(x_i) &\geq p_k f(x_k) + (1+p_k) f\left(\sum_{i=1}^{k-1} q_i f(x_i)\right) \\ &\geq f\left(p_k x_k + (1-p_k)\left(\sum_{i=1}^{k-1} q_i f(x_i)\right)\right) \\ &= f\left(\sum_{i=1}^{k} p_i x_i\right)\end{aligned}$$

□

Example 9.15. If $Y(x) = \ln x$ then because Y is concave in x then

$$\frac{1}{n}\sum_{i=1}^{n} \ln X_i \leq \ln \frac{\sum_{i=1}^{n} X_i}{n} \implies \sqrt[n]{\prod_{i=1}^{n} X_i} \leq \frac{1}{n}\sum_{i=1}^{n} X_i$$

This shows that the geometric mean is always less than the arithmetic mean.

Note 3. For the final exam, pay attention to the tutorial content on

1. Method of Moments for Gamma (Q1) where

   $$MM = \frac{2}{n}\sum_{i=1}^{n} x_i, ML = \max(X_1, ..., X_n)$$

   with $Var(MM) > Var(ML)$.

Index