ECON 5340 Class Notes
Review of Probability and Distribution Theory

1 Random Variables

Definition. Let c represent an element of the sample space C of a random experiment, $c \in C$. A random variable is a one-to-one function $X = X(c)$. An outcome of X is denoted x.

Example. Single Coin Toss

- $C = \{c = T; c = H\}$
- $X(c) = 0$ if $c = T$
- $X(c) = 1$ if $c = H$

1.1 Probability Distribution Function (pdf)

Two types:

1. Discrete pdf. A function $f(x)$ such that $f(x) \geq 0$, $\forall x$ and $\sum_x f(x) = 1$.
2. Continuous pdf. A function $f(x)$ such that $f(x) \geq 0$, $\forall x$ and $\int_{x=-\infty}^{\infty} f(x)dx = 1$.

Notes:

1. $\Pr(X = x) = f(x)$ in the discrete case, and $\Pr(X = x) = 0$ in the continuous case.
2. $\Pr(a \leq X \leq b) = \int_{x=a}^{b} f(x)dx$.

1.2 Cumulative Distribution Function (cdf)

Two types:

1. Discrete cdf. A function $F(x)$ such that $\sum_{X \leq x} f(x) = F(x)$.
2. Continuous cdf. A function $F(x)$ such that $\int_{-\infty}^{x} f(t)dt = F(x)$.

Notes:

1. $F(b) - F(a) = \int_{\infty}^{b} f(t)dt - \int_{\infty}^{a} f(t)dt$ where $b \geq a$.

2. $0 \leq F(x) \leq 1$.

3. $\lim_{x \to -\infty} F(x) = 0$.

4. $\lim_{x \to +\infty} F(x) = 1$.

5. If $x > y$, $F(x) \geq F(y)$.

2 Mathematical Expectations

Consider the continuous case only.

2.1 Mean

Definition. The mean or expected value of $g(X)$ is given by

$$E[g(x)] = \int_x g(x)f(x)dx.$$

Notes:

1. $E(X) = \mu = \int_x xf(x)dx$ is called the mean of X or the "first moment of the distribution".

2. $E(\cdot)$ is a linear operator. Let $g(X) = a + bX$.

$$\begin{aligned} E[g(X)] &= \int_x (a+bx)f(x)dx = \int_x af(x)dx + \int_x bxf(x)dx \\ &= E(a) + E(bX) = a + bE(X). \end{aligned}$$

3. Other measures of central tendency: median, mode.

2.2 Variance

Definition. The variance of $g(X)$ is given by

$$Var[g(X)] = E[\{g(X) - E[g(X)]\}^2] = \int_x \{g(x) - E[g(x)]\}^2 f(x)dx.$$

Notes:

1. Let $g(X) = X$. We have

$$
\begin{aligned}
Var(X) &= \sigma^2 = \int_x (x-\mu)^2 f(x)dx = \int_x x^2 f(x)dx - 2\mu \int_x x f(x)dx + \mu^2 \int_x f(x)dx \\
&= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2.
\end{aligned}
$$

2. $Var(X)$ is NOT a linear operator. Let $g(x) = a + bX$.

$$
Var[g(X)] = \int_x \{g(x) - g(\mu)\}^2 f(x)dx = \int_x b^2 (x-\mu)^2 f(x)dx = b^2 Var(X) = b^2\sigma^2.
$$

3. σ is called the standard deviation of X.

2.3 Other Moments

The measure $E(X^r)$ is called the "r^{th} moment of the distribution" while $E[(X-\mu)^r]$ is called the "r^{th} central moment of the distribution".

r	Central Moment	Measure
1	$E[(X-\mu)] = 0$	
2	$E[(X-\mu)^2] = \sigma^2$	variance (dispersion)
3	$E[(X-\mu)^3]$	skewness (asymmetry)
4	$E[(X-\mu)^4]$	kurtosis (tail thickness).

Moment Generating Function (MGF). The MGF uniquely determines a pdf when it exists and is given by

$$
M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx.
$$

The r^{th} moment of a distribution is given by

$$
\frac{d^r M(t)}{dt^r}|_{t=0}.
$$

2.4 Chebyshev's Inequality

Definition. Let X be a random variable with $\sigma^2 < \infty$. For any $k > 0$,

$$
\Pr(\mu - k\sigma \le X \le \mu + k\sigma) \ge 1 - \frac{1}{k^2}.
$$

Chebyshev's inequality is used to calculate upper (and lower) bounds on a random variable without having to know the exact distribution.

Example. Let $X \sim f(x)$ where

$$f(x) = \frac{1}{2\sqrt{3}}, \quad -\sqrt{3} < x < \sqrt{3}$$

and zero elsewhere. If we let $k = 3/2$, we get

$$\begin{aligned} Cheb &: \ \Pr(-3/2 \leq X \leq 3/2) \geq 1 - \frac{1}{(3/2)^2} = 5/9 = 0.\overline{55} \\ Exact &: \ \Pr(-3/2 \leq X \leq 3/2) = \int_{-3/2}^{3/2} \frac{1}{2\sqrt{3}} dx = \frac{1}{2\sqrt{3}}[(3/2) - (-3/2)] \simeq 0.866. \end{aligned}$$

3 Specific Probability Distributions

3.1 Normal pdf

If X has a normal distribution, then

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

where $-\infty < x < \infty$. In short-hand notation, $X \sim N(\mu, \sigma^2)$.

Notes:

1. The normal pdf is symmetric.

2. $Z = (X - \mu)/\sigma \sim N(0, 1)$ is called a standardized random variable and

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-0.5z^2)$$

is called the standard normal distribution.

3. Linear transformations of normal random variables are normal. If $Y = a + bX$ where $X \sim N(\mu, \sigma^2)$, then $Y \sim N(a + b\mu, b^2\sigma^2)$.

3.2 Chi-square pdf

If Z_i, $i = 1, ..., n$, are independently distributed $N(0,1)$ random variables,

$$Y = \sum\nolimits_{i=1}^{n} Z_i^2 \sim \chi^2(n)$$

where $E(Y) = n$ and $Var(Y) = 2n$.

<u>Exercise</u>. Find the MGF for $Y = Z^2$ and use it to derive the mean and variance.

<u>Answer</u>. We begin by calculating the MGF for Z^2 where $t < 0.5$:

$$M(t) = E(e^{tZ^2}) = \int_{-\infty}^{\infty} e^{tz^2} \phi(z) dz = \int_{-\infty}^{\infty} (2\pi)^{-0.5} e^{(t-0.5)z^2} dz = \int_{-\infty}^{\infty} (2\pi)^{-0.5} e^{-0.5(1-2t)z^2} dz.$$

Now using the method of substitution, let $w = \sqrt{(1-2t)}z$ so that $dw = (1-2t)^{1/2} dz$. Now making the substitution produces

$$M(t) = (1-2t)^{-1/2} \int_{-\infty}^{\infty} (2\pi)^{-0.5} e^{-0.5w^2} dw = (1-2t)^{-1/2}.$$

To calculate the mean, we take the first derivative of $M(t)$ and evaluate at $t = 0$:

$$\mu = \frac{dM(t)}{dt}|_{t=0} = (1-2t)^{-3/2}|_{t=0} = 1.$$

To calculate the variance, we take the second derivative of $M(t)$, evaluate at $t = 0$, and subtract μ^2:

$$\sigma^2 = \left[\frac{d^2 M(t)}{dt^2}|_{t=0} \right] - \mu^2 = 3(1-2t)|_{t=0} - \mu^2 = 2.$$

3.3 F pdf

If X_1 and X_2 are independently distributed $\chi^2(n_i)$ random variables,

$$F = \frac{X_1/n_1}{X_2/n_2} \sim F(n_1, n_2).$$

3.4 Student's t pdf

If $Z \sim N(0,1)$ and $X \sim \chi^2(n)$ are independent,

$$T = \frac{Z}{\sqrt{X/n}} \sim t(n).$$

3.5 Lognormal pdf

If $X \sim N(\mu, \sigma^2)$ then $Y = \exp(X)$ has the distribution

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp[-0.5(\frac{\ln(y) - \mu}{\sigma})^2]$$

for $y \geq 0$. Sometimes this is written as $y \sim LN(\mu, \sigma^2)$. The mean and variance of Y are $E(Y) = \exp(\mu + \sigma^2/2)$ and $Var(Y) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$.

Notes:

1. If $Y_1 \sim LN(\mu_1, \sigma_1^2)$ and $Y_2 \sim LN(\mu_2, \sigma_2^2)$ are independent random variables, then

$$Y_1 Y_2 \sim LN(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

3.6 Gamma pdf

The gamma distribution is given by

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-x/\beta)$$

for $0 \leq x < \infty$. The mean and variance are $E(X) = \alpha\beta$ and $Var(X) = \alpha\beta^2$.

Notes:

1. $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} \exp(-y) dy$ is called the gamma function, $\alpha > 0$.
2. $\Gamma(\alpha) = (\alpha - 1)!$ if α is a positive integer.
3. Greene sets $\beta = 1/\lambda$ and $\alpha = P$.
4. When $\alpha = 1$, you get the exponential pdf.
5. When $\alpha = n/2$ and $\beta = 2$, you get the chi-square pdf.

Example. Gamma distributions are sometimes used to model "waiting times". Let W be the waiting time until death for a human. Let $W \sim Gamma(\alpha = 1, \beta = 80)$ so that the expected waiting time until death is 80 years. (Note: $W \sim Exponential(\beta)$). Find the $\Pr(W \leq 30)$.

$$\begin{aligned}
\Pr(W &\leq 30) = \int_0^{30} \frac{1}{\Gamma(1)80} \exp(-w/80) dw = \frac{1}{80} \int_0^{30} \exp(-w/80) dw \\
&= \frac{1}{80} \left(-80 \exp(-w/80)\right)\Big|_{w=0}^{30} = -\left[\exp(-3/8) - \exp(0)\right] = 1 - 0.687 = 0.313.
\end{aligned}$$

3.7 Beta pdf

If X_1 and X_2 are independently distributed Gamma random variables then $Y_1 = X_1 + X_2$ and $Y_2 = X_1/Y_1$ are independently distributed. The marginal distribution $f_2(y_2)$ of $f(y_1, y_2)$ is called the beta pdf:

$$g(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (y/c)^{\alpha - 1} \left[1 - (y/c)\right]^{\beta - 1} (1/c)$$

where $0 \leq y \leq c$. The mean and variance are $E(Y) = c\alpha/(\alpha + \beta)$ and $Var(Y) = c^2\alpha\beta/(\alpha + \beta + 1)$.

3.8 Logistic pdf

The logistic distribution is

$$f(x) = \Lambda(x) \left[1 - \Lambda(x)\right]$$

where $-\infty < x < \infty$ and $\Lambda(x) = (1 + \exp(-x))^{-1}$. The mean and variance are $E(X) = 0$ and $Var(X) = \pi^2/3$. A useful property of the logistic distribution is that the cdf has a closed-form solution

$$F(x) = \Lambda(x).$$

3.9 Cauchy pdf

If X_1 and X_2 are independently distributed $N(0, 1)$, then

$$Y = X_1/X_2 \sim f(y) = \frac{1}{\pi(1 + y^2)}$$

where $-\infty < y < \infty$. The mean and the variance of the Cauchy pdf do not exist because the tails are too thick.

3.10 Binomial pdf

The distribution for x successes in n trials is

$$b(n, \alpha, x) = \binom{n}{x} \alpha^x (1-\alpha)^{n-x}$$

where $x = 0, 1, ..., n$ and $0 \leq \alpha \leq 1$. The mean and variance of the binomial distribution are $E(X) = n\alpha$ and $Var(X) = n\alpha(1-\alpha)$. The combinatorial formula for the number of ways to choose x objects from a set n distinct objects is

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

3.11 Poisson pdf

The Poisson pdf is often used to model the number of changes in a fixed interval. The Poisson pdf is

$$f(x) = \frac{\exp(-\lambda)\lambda^x}{x!}$$

where $x = 0, 1, ...$ and $\lambda > 0$. The mean and variance are $E(X) = Var(X) = \lambda$.

Example. Let the probability of Spina Bifida in one child be 1/1000. Let X be the number of Spina Bifida cases in 3000 births. With stochastic independence and $X \sim Poisson(\lambda = 3000/1000 = 3)$, find the probability that there are five cases in 3000 births.

Answer.

$$\Pr(X = 5) = \frac{\exp(-3)3^5}{5!} = 0.101.$$

4 Distributions of Functions of Random Variables

Let X_1, $X_2, ..., X_n$ have joint pdf $f(x_1, ..., x_n)$. What is the distribution of $Y = g(X_1, X_2, ..., X_n)$? To answer this question, we will use the change-of-variable technique.

Change of Variable Technique. Let X_1 and X_2 have joint pdf $f(x_1, x_2)$. Let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ be the transformed random variables. If A is the set where $f > 0$, then let B be the set defined by the one-to-one transformation of A to B. Then

$$g(y_1, y_2) = f(h_1(y_1, y_2), h_2(y_1, y_2)) \cdot |J|$$

where $(y_1, y_2) \in B$, $x_1 = h_1(y_1, y_2)$, $x_2 = h_2(y_1, y_2)$ and

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}.$$

Example. Let X_1 and X_2 be uniformly distributed on $0 \leq X_i \leq 1$. The random sample X_1, X_2 is jointly distributed

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) = 1$$

over $0 \leq x_1, x_2 \leq 1$ and zero elsewhere. Find the joint distribution of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.

Answer. We know that $x_1 = h_1(y_1, y_2) = 0.5(y_1 + y_2)$ and $x_2 = h_2(y_1, y_2) = 0.5(y_1 - y_2)$. We also know that

$$J = \begin{vmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \end{vmatrix} = -0.5.$$

Therefore,

$$g(y_1, y_2) = f_1(h_1(y_1, y_2)) f_2(h_1(y_1, y_2)) \cdot |J| = 0.5$$

where $(y_1, y_2) \in B$ and zero elsewhere.

5 Joint Distributions

5.1 Joint pdfs and cdfs

- A joint pdf for X_1 and X_2 gives $\Pr(X_1 = x_1, X_2 = x_2) = f(x_1, x_2)$. A proper joint pdf will have the property $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 dx_1 = 1$ and $f(x_1, x_2) \geq 0$ for all x_1 and x_2.

- A joint cdf for X_1 and X_2 is $\Pr(X_1 \leq x_1, X_2 \leq x_2) = F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(t_1, t_2) dt_2 dt_1$.

5.2 Marginal Distributions

The marginal pdf of X_1 is found by integrating over all X_2:

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$$

and likewise for X_2.

<u>Example</u>. Let X_1 and X_2 have joint pdf

$$f(x_1, x_2) = 2,\ 0 < x_1 < x_2 < 1$$

and zero elsewhere. Is this a proper pdf?

$$\int_0^1 \int_{x_1}^1 2dx_2dx_1 = \int_0^1 \left[2x_2|_{x_2=x_1}^1\right] dx_1 = \int_0^1 2(1-x_1)dx_1 = 2x_1|_{x_1=0}^1 - x_1^2|_{x_1=0}^1 = 2 - 1 = 1.$$

So yes, this is a proper pdf. The marginal distribution for X_1 is

$$f_1(x_1) = \int_{x_1}^1 2dx_2 = 2x_2|_{x_2=x_1}^1 = 2(1-x_1),\ 0 < x_1 < 1$$

and zero elsewhere. The marginal distribution for X_2 is

$$f_2(x_2) = \int_0^{x_2} 2dx_1 = 2x_1|_{x_1=0}^{x_2} = 2x_2,\ 0 < x_2 < 1$$

and zero elsewhere.

Notes:

1. Two random variables are stochastically independent if and only if $f_1(x_1)f_2(x_2) = f(x_1, x_2)$.

2. In our example, X_1 and X_2 are not independent because $f_1(x_1)f_2(x_2) = 4x_2 - 4x_1x_2 \neq 2 = f(x_1, x_2)$.

3. Moments (e.g., means and variances) in joint distributions are calculated using marginal densities (e.g., $E(X_1) = \int x_1 f_1(x_1) dx_1$.

5.3 Covariance and Correlation

Definition. The covariance between X and Y is

$$cov(X,Y) = E\left[(X-\mu_x)(Y-\mu_y)\right] = E(XY) - \mu_x\mu_y.$$

Definition. The correlation coefficient between X and Y removes the dependence on the unit of measurement:

$$\rho = corr(X,Y) = \frac{cov(X,Y)}{\sigma_x\sigma_y}$$

where $-1 \le \rho \le 1$.

Notes:

1. If X and Y are independent, then $cov(X,Y) = 0$:

$$\begin{aligned} cov(X,Y) &= E(XY) - \mu_x\mu_y = \int\int xyf_x(x)f_y(y)dydx - \mu_x\mu_y \\ &= \int xf_x(x)dx \int yf_y(y)dy - \mu_x\mu_y = \mu_x\mu_y - \mu_x\mu_y = 0. \end{aligned}$$

2. However, $cov(X,Y) = 0$ does not imply stochastic independence. Consider the following joint distribution table

		y			$f_x(x)$
		-1	0	1	
	-1	0	0	1/3	1/3
x	0	0	1/3	0	1/3
	1	0	0	1/3	1/3
$f_y(y)$		0	1/3	2/3	

where $\mu_x = 0$, $\mu_y = 2/3$ and

$$\begin{aligned} cov(X,Y) &= \sum\sum(x-\mu_x)(y-\mu_y)f(x,y) \\ &= (-1)(1/3)(1/3) + (0)(-2/3)(1/3) + (1)(1/3)(1/3) = 0. \end{aligned}$$

However, X and Y are not independent because for $(x, y) = (0, 0)$ we have

$$f_x(0)f_y(0) = 1/9 \neq f(0, 0) = 1/3.$$

6 Conditional Distributions

Definition. The conditional pdf for X given Y is

$$f(x|y) = \frac{f(x, y)}{f_y(y)}.$$

Notes:

1. If X and Y are independent, $f(x|y) = f_x(x)$ and $f(y|x) = f_y(y)$.
2. The conditional mean is $E(X|Y) = \int xf(x|y)dx = \mu_{x|y}$.
3. The conditional variance is $Var(X|Y) = \int (x - \mu_{x|y})^2 f(x|y)dx$.

7 Multivariate Distributions

Let $X = (X_1, ..., X_n)'$ be a $(n \times 1)$ column vector of random variables. The mean and variance of X is

$$\mu = E(X) = (\mu_1, ..., \mu_n)'$$

and

$$\Sigma = Var(X) = E[(X - \mu)(X - \mu)'] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2n} \\ \vdots & & \ddots & \\ \sigma_{n1} & \sigma_{n2} & & \sigma_{nn} \end{bmatrix}.$$

Notes:

1. Let $W = A + BX$. Then $E(W) = A + BE(X)$.

2. The variance of W is

$$\begin{aligned} Var(W) &= E[(W-E(W))(W-E(W))'] = E[(BX-BE(X))(BX-BE(X))'] \\ &= E[B(X-E(X))(X-E(X))'B'] = B\Sigma B'. \end{aligned}$$

7.1 Multivariate Normal Distributions

Let $X = (X_1, ..., X_n)' \sim N(\mu, \Sigma)$. The form of the multivariate normal pdf is

$$f(x) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp[-0.5(x-\mu)'\Sigma^{-1}(x-\mu)].$$

See Gauss example P.1 for an example of a bivariate normal density function.

7.2 Quadratic Form in a Normal Vector

If $(X-\mu)$ is a normal vector, then the quadratic form $Q = (X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2(n)$.

Proof. The moment generating function of Q is

$$\begin{aligned} M(t) &= E(e^{tQ}) \\ &= \int \cdots \int (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp[t(x-\mu)'\Sigma^{-1}(x-\mu) - 0.5(x-\mu)'\Sigma^{-1}(x-\mu)] dx_1 \cdots dx_n \\ &= \int \cdots \int (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp[-0.5(x-\mu)'(1-2t)\Sigma^{-1}(x-\mu)] dx_1 \cdots dx_n. \end{aligned}$$

Next, multiply and divide by $(1-2t)^{n/2}$:

$$\begin{aligned} M(t) &= \frac{\int \cdots \int (2\pi)^{-n/2} |\Sigma/(1-2t)|^{-1/2} \exp[-0.5(x-\mu)'(1-2t)\Sigma^{-1}(x-\mu)] dx_1 \cdots dx_n}{(1-2t)^{n/2}} \\ &= (1-2t)^{-n/2}, \ t < 0.5. \end{aligned}$$

The numerator is the integral of a multivariate normal random distribution with variance $\Sigma/(1-2t)$ and so it equals one. $M(t)$ then simplifies to the MGF for a $\chi^2(n)$ random variable.

7.3 A Couple of Important Theorems

1. Let $X \sim N(0, I)$ and $A^2 = A$ (i.e., A is idempotent). $X'AX \sim \chi^2(r)$ where the rank of A is r.

2. Let $X \sim N(0, I)$. $X'AX$ and $X'BX$ are stochastically independent iff $A \cdot B = 0$.