APPM 5520: Introduction to Mathematical Statistics
Course Notes

Robert Lindsey

Fall 2009

1 Week 1

1.1 Lecture 1: 8-24-09

The following notation will be used throughout the course:

- Probability density function (pdf): f
- Cumulative distribution function (cdf): F

1.1.1 Basic Definitions

When we say that a random variable X has a "probability density function," we mean one of two things:

1. If X is discrete, then X has a probability *mass* function $f(x)$, where $f(x) = P(X = x)$. The probability mass function tells us the likelihood of the random variable X taking on the value x.
2. If X is continuous, then the probability density function is defined differently out of necessity. Because $P(X = x) = 0$ for all x, it makes no sense to define $f(x)$ as we did in the discrete case. Consequently, we are interested in $P(a < X \leq b)$, the probability that X falls within the range $[a,b)$. Thus, we define a probability density function $f(x)$ for continuous random variables in terms area: $P(a < X \leq b) = \int_a^b f(x)dx$.

Whether X is discrete or continuous, the cumulative distribution function is defined as $F(x) = P(X \leq x)$ (the probability that X takes on a value less than or equal to x).

- If X is continuous, $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$
- If X is discrete, $F(x) = \sum_{u=-\infty}^{x} P(X = u)$

A probability density function is the derivative of a corresponding cumulative distribution function, meaning $f(x) = \frac{d}{dx}F(x)$. Note that if X is discrete and $x \in \mathbb{Z}$, $f(x) = F(x) - F(x-1)$.

1.1.2 The Geometric Distribution

Consider a sequence of independent trials in an experiment where each trial can be a "success" or "failure" (abbreviated S or F). As a side note, the sequence of trials is often called a *Bernoulli process* and each trial in the sequence is called a *Bernoulli trial*. In each trial, let the probability of success be p, where $0 \leq p \leq 1$. Let X be the number of trials till the first success. The following is the probability mass function of X:

$$\begin{aligned}
P(X=1) &= p \\
P(X=2) &= (1-p)p \\
&\vdots \\
P(X=x) &= \begin{cases} (1-p)^{x-1}p & \text{for x = } 1,2,3\ldots \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

Intuitively, $(1-p)$ is the probability of failure on a trial, so $(1-p)^{x-1}$ is the probability of $x-1$ consecutive failures. Thus $(1-p)^{x-1}p$ is the probability of $x-1$ consecutive failures followed by a success. This probability mass function is called the **geometric distribution**.

Some textbooks define the geometric distribution in terms of the number of failures before the first success. Under this assumption, we have $P(X=x)=(1-p)^x p$ where $x=0,1,2\ldots$. The range of x starts at 0 because it is possible to succeed on the first trial, hence have no failures.

We usually say that $X \sim \text{geom}(p)$, where the tilda means "has the distribution."

1.1.3 The Exponential Distribution

Imagine sitting at the door to a bank. Assume that the arrival rate of people is 4.3 customers per minute. Also assume that the number of arrivals in 2 non-overlapping periods of time are independent. Let X be the time between any two consecutive arrivals. We can (but won't) show that $f(x)=4.3e^{-4.3x}$, where $x>0$. X is said to have the **exponential distribution** with rate 4.3, $X \sim \exp(\text{rate}=4.3)$. The mean inter-arrival time is the inverse of the arrival rate.

1.1.4 Indicator Notation

Let A be a set. The **indicator function** I for A is defined as:

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Example: Let $X \sim geom(p)$. The probability mass function written in indicator notation is $f(x)=(1-p)^{x-1}p\mathrm{I}_{\{0,1,\ldots\}}(x)$

1.2 Lecture 2: 8-26-09

1.2.1 Marginal Probability Density Functions

Suppose that X and Y are continuous random variables with joint probability distribution function $f(x,y)$. The **marginal** probability distribution function for X is $f_x(x)=\int_{-\infty}^{\infty} f(x,y)dy$. Similarly, the marginal for Y is $f_y(y)=$

$\int_{-\infty}^{\infty} f(x,y)dx$. In both cases, we are integrating with respect to the variable we want to get rid of. Although we are integrating from negative to positive infinity, the pdf's may only have a small support (part of domain where the function is non-zero).

1.2.2 Independence

X and Y are **independent** if and only if $f(x,y) = f_x(x)f_y(y)$

Example: Let X and Y have joint pdf $f(x,y) = xy$, where $0 < x < 1$ and $0 < y < 2$. Implicitly, f is 0 outside of this range. The marginals of X and Y are:

$$\begin{aligned}
f_x(x) &= \int_{-\infty}^{\infty} f(x,y)dy = \int_{0}^{2} (xy)\,dy = \left(.5xy^2\right)|_0^2 = 2x \quad \text{where } 0 < x < 1 \\
f_y(y) &= \int_{-\infty}^{\infty} f(x,y)dx = \int_{0}^{1} (xy)\,dx = \frac{1}{2}y \quad \text{where } 0 < y < 2 \\
f_x(x)f_y(y) &= f(x,y)
\end{aligned}$$

Therefore, X and Y are independent.

Example: Let $f(x,y) = 8xy$, where $0 < x < y < 1$.

$$\begin{aligned}
f_x(x) &= \int_{-\infty}^{\infty} f(x,y)dy = \int_{x}^{1} (8xy)\,dy = \left(4xy^2\right)|_x^1 = 4x\left(1-x^2\right) \quad \text{where } 0 < x < 1 \\
f_y(y) &= \int_{-\infty}^{\infty} f(x,y)dx = \int_{0}^{y} (8xy)\,dx = 4y^3 \\
f_x(x)f_y(y) &\neq f(x,y)
\end{aligned}$$

Therefore, X and Y are not independent.

It is not always the case that a messy Indicator Function (as in the above example) implies non-independence.

Consider $f(x,y) =$ stuff that factors $I_{(0,\infty)}(xy)\,I_{(0,\infty)}(y-xy)$. By graphing out where those two indicator functions are "on" (e.g. where is xy in the range $(0,\infty)$), we can simplify the indicator notation to $I_{(0,1)}(x)\,I_{(0,\infty)}(y)$. Since the stuff in front of the indicator notation factors (by assumption) and the indicator functions form a rectangular region, the X and Y are independent.

1.2.3 Finding Distributions of Transformations of Random Variables

Example: Suppose X is a discrete random variable, $X \sim \text{geom}(p)$. Find the pdf of $y = 5x$.

$$f_y(y) = P(Y=y) = P(5X=y) = P\left(X = \frac{y}{5}\right) = f_x\left(\frac{y}{5}\right) = (1-p)^{\frac{y}{5}-1} p\mathrm{I}_{\{1,2,\dots\}}\left(\frac{y}{5}\right) = (1-p)^{\frac{y}{5}-1} p\mathrm{I}_{\{5,10,\dots\}}(y)$$

The above approach only works in the discrete case, since $P(X = x) = 0$ for a continuous random variable. To find the distribution of a transformation of a continuous random variable, you need to work with the cumulative distribution function and then take its derivative (because $f_y(y) = \frac{d}{dy}F_y(y)$).

Example: Assume X has pdf $f_x(x)$ and cdf $F_x(x)$ and that $Y = g(X)$. g is our transformation function. Although it isn't always going to be true, for now, assume g is invertible. We know the following:

$$g \text{ is increasing} \leftrightarrow g^{-1} \text{ is increasing}$$
$$g \text{ is decreasing} \leftrightarrow g^{-1} \text{ is decreasing}$$

How do we write the pdf for y? We must evaluate the increasing/decreasing cases of g separately (ultimately they'll be proven to be the same).

Case 1: g is increasing: $F_y(y) = P(Y \leq y) = P(g(x) \leq y) = P(g^{-1}(g(x)) \leq g^{-1}) = P(x \leq g^{-1}(y)) = F_x(g^{-1}(y))$. This just means that the cdf of Y is the cdf of X transformed by $g^{-1}(y)$. Because of this, we can write the pdf of y in terms of the pdf of x:

$$f_y(y) = \frac{d}{dy}F_y(y) = \frac{d}{dy}F_x\left(g^{-1}(y)\right) = F_x'\left(g^{-1}(y)\right)\frac{d}{dy}g^{-1}(y) = f_x\left(g^{-1}(y)\right)\frac{d}{dy}g^{-1}(y)$$

Case 2: g is increasing: We ran out of time in class, but just assert for now that $f_y(y) = f_x(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$

1.3 Lecture 3: 8-28-09

1.3.1 Integration

In this class, we rarely need to do integration. What integration we have to do can often be rewritten as a known distribution (hence with an integral of 1).

Example:

$$\int_0^\infty e^{-x^2}dx = \frac{1}{2}\int_{-\infty}^\infty e^{-x^2}dx$$

This looks sort-of like a normal distribution with mean 0 and variance .5. Let's now rewrite it as a normal distribution and solve it:

$$\frac{1}{2}\sqrt{2\pi\frac{1}{2}}\int_{-\infty}^\infty \left(2\pi\frac{1}{2}\right)^{-\frac{1}{2}} e^{-x^2}dx = \frac{1}{2}\sqrt{\pi}\int_{-\infty}^\infty N\left(0,\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}$$

1.3.2 Finding Distributions of Transformations of Random Variables (cont.)

Last lecture, we did not finish showing that when g is decreasing, we get the same result as if it were increasing.

Because g is decreasing, we know that $F_y(y) = P(Y \leq y) = P(g(x) \leq y)$. Because g^{-1} is also decreasing and probabilities sum to 1, we know that $P(x \geq g^{-1}(y)) = 1 - P(x < g^{-1}(y))$.

So, the pdf is given by:

$$f_y(y) = \frac{d}{dy}F_y(y) = \frac{d}{dy}\left[1 - F_x(g^{-1}(y))\right] = 0 - F_x'(g^{-1}(y))\frac{d}{dy}g^{-1}(y) = -f_x(g^{-1}(y))\frac{d}{dy}g^{-1}(y)$$

Since $g^{-1}(y) < 0$, the above quantity is positive[1]. Thus we can conclude that whether g is increasing or decreasing, the pdf after the transformation g is:

$$f_y(y) = f_x(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$$

1.3.3 Gamma Distribution

Suppose that X has a **gamma distribution** with parameters α and β. We write $X \sim \Gamma(\alpha, \beta)$. The pdf for a gamma distribution is:

$$f_x(x) = \frac{1}{\Gamma(\alpha)}\beta^\alpha x^{\alpha-1}e^{-\beta x}I_{(0,\infty)}(x)$$

Example similar to homework 1, problems 2 and 3:

Suppose $Y = 5X$. Find the distribution and name it. Solution: We have $y = g(x) = 5x \Rightarrow x = g^{-1}(y) = \frac{y}{5}$. We find that:

$$f_y(y) = f_x(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| = \left(\frac{1}{\Gamma(\alpha)}\beta^\alpha\left(\frac{y}{5}\right)^{\alpha-1}e^{-\frac{\beta y}{5}}I_{(0,\infty)}\left(\frac{y}{5}\right)\right)\frac{1}{5}$$

$$f_y(y) = \frac{1}{\Gamma(\alpha)}\left(\frac{\beta}{5}\right)^\alpha y^{\alpha-1}e^{-\frac{\beta y}{5}}I_{(0,\infty)}(y)$$

Therefore, $y \sim \Gamma\left(\alpha, \frac{\beta}{5}\right)$.

1.3.4 Gamma Function

What is this mysterious $\Gamma(\alpha)$ we used in defining the gamma distribution? It is given by:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$$

Some basic properties of the gamma function:

1. $\Gamma(1) = 1$
2. $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$.
3. If n is a positive integer, $\Gamma(n) = (n-1)!$

[1] How do we know g-inverse is negative??

1.3.5 Bivariate to Bivariate Transformations

Given X_1 and X_2 with joint probability density function $f_{x_1 x_2}(x_1, x_2)$, we want to go to Y_1 and Y_2 with joint probability density function $f_{y_1 y_2}(y_1, y_2)$, where $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$. This is like what we did previously, but for two random variables.

The analog of $\frac{d}{dy} g^{-1}(y)$ is the (backward) Jacobian of the transformation:

$$\mathbf{J_B} = \begin{vmatrix} \frac{dx_1}{dy_1} & \frac{dx_1}{dy_2} \\ \frac{dx_2}{dy_1} & \frac{dx_2}{dy_2} \end{vmatrix}$$

The vertical bars indicate that we're taking the determinant of the matrix of partial derivatives.

2 Week 2

2.1 Lecture 4: 8-31-09

2.1.1 Bivariate to Bivariate Transformations

This is picking up from where we left off last time.

We have:

$$f_{y_1 y_2}(y_1, y_2) = f_{x_1 x_2}\left(g_1^{-1}(y_1, y_2), g_1^{-1}(y_1, y_2)\right) |\mathbf{J_B}|$$

Note that $|\mathbf{J_B}|$ is the absolute value of the backward Jacobian. The backward Jacobian is just a number. It is a function of y_1 and y_2 (a forward Jacobian would have been a function of x_1 and x_2).

2.1.2 A Very Long Example

- Let $X_1, X_2 \sim \Gamma(\alpha, \beta)$ where X_1 and X_2 are identically and independently distributed (iid).
- Find the distribution of $Y = \frac{X_1}{X_1 + X_2}$.

To do this define, $y_1 = \frac{x_1}{x_1 + x_2}$ and $y_2 =$ anything. We'll find $f_{y_1 y_2}$ and marginalize out y_2:

$$f_{y_1}(y_1) = \int_{-\infty}^{\infty} f_{y_1 y_2}(y_1, y_2)\, dy_2$$

For convenience, we'll set $y_2 =$ the denominator of Y $= x_1 + x_2$. Since x_1 and x_2 are independent,

$$
\begin{aligned}
f_{x_1x_2}(x_1,x_2) &= f_{x_1}(x_1)\, f_{x_2}(x_2) \\
&= \left(\frac{1}{\Gamma(\alpha)}\beta^{\alpha}x_1^{\alpha-1}e^{-\beta x_1}I_{(0,\infty)}(x_1)\right)\left(\frac{1}{\Gamma(\alpha)}\beta^{\alpha}x_2^{\alpha-1}e^{-\beta x_2}I_{(0,\infty)}(x_2)\right) \\
&= \frac{1}{(\Gamma(\alpha))^2}\beta^{2\alpha}(x_1x_2)^{\alpha-1}e^{-\beta(x_1+x_2)}I_{(0,\infty)}(x_1)\, I_{(0,\infty)}(x_2)
\end{aligned}
$$

Now we write x_1 and x_2 in terms of y_1 and y_2. In other words, we are finding $g_1^{-1}(y_1,y_2)$ and $g_2^{-1}(y_1,y_2)$.

$$
y_1 = g_1(x_1,x_2) = \frac{x_1}{x_1+x_2}
$$

$$
y_2 = g_2(x_1,x_2) = x_1 + x_2
$$

$$
y_1 = \frac{x_1}{y_2}
$$

$$
x_1 = y_1y_2 = g_1^{-1}(y_1,y_2)
$$

$$
x_2 = y_2 - x_1 = y_2 - y_1y_2 = g_2^{-1}(y_1,y_2)
$$

Since $x_1 = y_1y_2$ and $x_2 = y_2 - y_1y_2$, the backward Jacobian is given by:

$$
\mathbf{J_B} = \begin{vmatrix} \frac{dx_1}{dy_1} & \frac{dx_1}{dy_2} \\ \frac{dx_2}{dy_1} & \frac{dx_2}{dy_2} \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ -y_2 & (1-y_1) \end{vmatrix} = y_2(1-y_1) - (-y_1y_2) = y_2
$$

Using the expression we found earlier for $f_{x_1x_2}(x_1,x_2)$ and also the previously derived fact that

$$
f_{y_1y_2}(y_1,y_2) = f_{x_1x_2}\left(g_1^{-1}(y_1,y_2), g_1^{-1}(y_1,y_2)\right)|\mathbf{J_B}|
$$

We can now say that

$$
\begin{aligned}
f_{y_1y_2}(y_1,y_2) &= \frac{1}{(\Gamma(\alpha))^2}\beta^{2\alpha}(y_1y_2(y_2-y_1y_2))^{\alpha-1}e^{-\beta(y_1y_2+y_2-y_1y_2)}I_{(0,\infty)}(y_1y_2)\, I_{(0,\infty)}(y_2-y_1y_2)\,|y_2| \\
&= |y_2|\frac{1}{(\Gamma(\alpha))^2}\beta^{2\alpha}(y_2)^{2\alpha-2}[y_1(1-y_1)]^{\alpha-1}e^{-\beta y_2}I_{(0,1)}(y_1)\, I_{(0,\infty)}(y_2) \\
&= \frac{1}{(\Gamma(\alpha))^2}\beta^{2\alpha}(y_2)^{2\alpha-1}[y_1(1-y_1)]^{\alpha-1}e^{-\beta y_2}I_{(0,1)}(y_1)\, I_{(0,\infty)}(y_2)
\end{aligned}
$$

In that last step, we moved the $|y|$ into the $y_2^{2\alpha-2}$ term. Because of the indicator function, we do not need to worry about y_2 being negative and can drop the absolute value. Now we can finally marginalize out y_2 from the joint probability density function. This is done as follows:

$$
\begin{aligned}
f_{y_1}(y_1) &= \int_{-\infty}^{\infty} f_{y_1 y_2}(y_1, y_2)\, dy_2 \\
&= \int_{0}^{\infty} \frac{1}{(\Gamma(\alpha))^2} \beta^{2\alpha} y_2^{2\alpha-1} \left[y_1(1-y_1)\right]^{\alpha-1} e^{-\beta y_2} I_{(0,1)}(y_1)\ dy_2 \\
&= \tfrac{1}{(\Gamma(\alpha))^2} \beta^{2\alpha} I_{(0,1)}(y_1) \left[y_1(1-y_1)\right]^{\alpha-1} \int_{0}^{\infty} y_2^{2\alpha-1} e^{-\beta y_2}\ dy_2
\end{aligned}
$$

Note that the indicator function for y_2 is no longer needed, since we are integrating over its support. Also note that the integrand is basically $\Gamma(2\alpha, \beta)$ sans some constants. Our job now is to give it those constants so that we don't have to do a messy integration. We will multiply the outside of the integral by $\Gamma(2\alpha)$ and divide the inside by the inverse of that. We also will move $\beta^{2\alpha}$ inside the integrand.

$$
\begin{aligned}
f_{y_1}(y_1) &= \tfrac{\Gamma(2\alpha)}{(\Gamma(\alpha))^2} I_{(0,1)}(y_1) \left[y_1(1-y_1)\right]^{\alpha-1} \int_{0}^{\infty} \frac{1}{\Gamma(2\alpha)} \beta^{2\alpha} y_2^{2\alpha-1} e^{-\beta y_2}\ dy_2 \\
&= \tfrac{\Gamma(2\alpha)}{(\Gamma(\alpha))^2} I_{(0,1)}(y_1) \left[y_1(1-y_1)\right]^{\alpha-1} \int_{0}^{\infty} \Gamma(2\alpha, \beta) \\
&= \tfrac{\Gamma(2\alpha)}{(\Gamma(\alpha))^2} \left[y_1(1-y_1)\right]^{\alpha-1} I_{(0,1)}(y_1) \\
&= \tfrac{\Gamma(2\alpha)}{(\Gamma(\alpha))^2} y_1^{\alpha-1} (1-y_1)^{\alpha-1} I_{(0,1)}(y_1) \\
Y &\sim \beta(\alpha, alpha)
\end{aligned}
$$