Math 502AB - Lecture 12

Dr. Jamshidian

October 4, 2017

1 Lecture - Part 1

1.1 Independent Events

1.1.1 Theorem:

If $X \perp\!\!\!\perp Y$, then for functions g and h:

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

Proof:

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{X,Y}(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy \quad \text{by independence} \\ &= \left(\int_{-\infty}^{\infty} g(x)f_X(x)dx\right)\left(\int_{-\infty}^{\infty} h(y)f_Y(y)dy\right) \\ &= E[g(X)]E[h(Y)] \end{aligned}$$

1.1.2 Theorem:

Let $X \perp\!\!\!\perp Y$. Then for any $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Proof:

Let

$$g(x) = \mathcal{I}_{\{X \in A\}} \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

And, similarly, define $h(Y) = \mathcal{I}_{\{y \in B\}}$. Note that, since an indicator function is simply a *Bernoulli* random variable, we then have:

$$\begin{aligned} P(X \in A, Y \in B) &= E[g(X)h(Y)] \\ &= E[g(X)]E[h(Y)] \\ &= P(X \in A)P(Y \in B) \end{aligned}$$

1.1.3 Theorem:

Let $X \perp\!\!\!\perp Y$, and $M_X(t)$ and $M_Y(t)$ be the *moment generating functions* of X and Y respectively. Then, the *mgf* of $Z = X + Y$ is:

$$M_Z(t) = M_X(t)M_Y(t)$$

Note: This theorem is extremely useful when trying to find the distribution of the sum of two random variables (provided that the *moment generating functions* exist for both variables).

Proof

$$\begin{aligned} M_Z(t) = E[e^{t(X+Y)}] &= E[e^{tX}e^{tY}] \\ &= E[e^{tX}]E[e^{tY}] \\ &= M_X(t)M_Y(t) \end{aligned}$$

Example:

Let $X \perp\!\!\!\perp Y$, and $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$. Show that

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ &= \exp\left(\mu_1 t + \frac{\sigma_1^2 t^2}{2}\right) \cdot \exp\left(\mu_2 t + \frac{\sigma_2^2 t^2}{2}\right) \\ &= \exp\left((\mu_1 + \mu_2)\,t + \frac{\sigma_1^2 + \sigma_2^2}{2} t^2\right) \end{aligned}$$

1.2 Bivariate Transformations

Suppose we have two random variables X and Y an we are interested in the distribution involving:

$$\begin{cases} U = g_1(x, y) \\ V = g_2(x, y) \end{cases}$$

- The first problem we will look at is a function of two random variables, which is a random variable itself (for example: $X + Y$, as in the above example).
- The second problem we will look at is, given U and V, we want to understand the joint distribution of these two random variables.

1.2.1 Problem 1:

Examples:

1. Let $X \perp\!\!\!\perp Y$, with $X \sim Poisson(\lambda)$ and $Y \sim Poisson(\theta)$. Obtain the distribution of $U = X + Y$.

   The easy way to approach this problem is by using the *moment generating function*. However, here we will look at the *pmf*:

$$\begin{aligned}
P(U = u) = P(X + Y = u) &= \sum_{y=0}^{\infty} P(X + Y = u | Y = y) P(Y = y) \\
&= \sum_{y=0}^{\infty} P(X = u - y) P(Y = y) \\
&= \sum_{y=0}^{u} P(X = u - y) P(Y = y) \\
&= \sum_{y=0}^{u} \frac{e^{-\lambda}\lambda^{u-y}}{(u-y)!} \cdot \frac{e^{-\theta}\theta^{y}}{y!} \\
&= \frac{e^{-(\lambda+\theta)}}{u!} \sum_{y=0}^{u} \frac{u!}{(u-y)!y!} \theta^{y}\lambda^{u-y} \\
&= \frac{e^{-(\lambda+\theta)}}{u!} (\theta + \lambda)^{u} \\
&\sim Poisson(\theta + \lambda)
\end{aligned}$$

2. Let $f(x, y) = xe^{-x(y+1)}$, $x > 0$, $y > 0$ be the joint density of X and Y. Find the *pdf* of $W = XY$

$$\begin{aligned}
F_W(w) = P(W \leq w) &= P(XY \leq w) \\
&= P\left(Y \leq \frac{w}{X}\right) \\
&= \int_0^{\infty} \int_0^{\frac{w}{x}} xe^{-x(y+1)} dydx \\
&= 1 - e^{-w} \\
w &\sim exp(1)
\end{aligned}$$

1.2.2 Problem 2:

Now, consider the case where X, Y are jointly distributed. Let $U = g_1(X, Y)$ and $V = g_2(X, Y)$, and assume that this transformation can be inverted, such that $X = h_1(U, V)$ and $Y = h_2(U, V)$. Furthermore, assume that g_1 and g_2 have continuous partial derivatives, and:

$$\mathcal{J}(x,y) = \det \begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix} \neq 0 \quad \forall x, y$$

Then:

$$f_{U,V}(u,v) = f_{X,Y}\left(h_1(u,v), h_2(u,v)\right) \left|\mathcal{J}^{-1}\left(h_1(u,v), h_2(u,v)\right)\right|$$

Examples:

1. Let X and Y be jointly distributed with joint pdf $f_{X,Y}$. Let $U = X + Y$ and $V = X - Y$. Obtain the joint pdf of (U, V). Solving for X and Y, we get $X = \frac{U+V}{2} = h_1(u,v)$ and $Y = \frac{U-V}{2} = h_2(u,v)$. Then, by the process above:

$$\mathcal{J}(x,y) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$
$$f_{U,V}(u,v) = f_{X,Y}\left(\frac{U+V}{2}, \frac{U-V}{2}\right) \cdot \frac{1}{2}$$

2. Let $X, Y \sim Unif(0,1)$ (**iid**). Obtain the joint pdf of $U = X + Y$ and $V = X - Y$.

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) = \begin{cases} 1 & 0 \leq x \leq 1;\ 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Thus, by the result of the previous example, we know that

$$f_{U,V}(u,v) = \frac{1}{2}$$

But over *what* range? By drawing a picture, we get:

$$f_{U,V}(u,v) = \begin{cases} \frac{1}{2} & 0 \leq u - v \leq 2;\ 0 \leq u + v \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Note: It is important to remember that when making a transformation, it is not only the *density* that is changing, but the *domain as well.*

3. Let $Z \sim N(0,1)$ and $X \sim \chi^2_{(\nu)}$. Let $Z \perp\!\!\!\perp X$. Obtain the distribution of

$$U = \frac{Z}{\sqrt{X/\nu}}$$

If you have fractions of two random variables, you might want to define a new *random variable* equal to the denominator:

$$V = \sqrt{X/\nu}$$

Note:

$$f_{Z,X}(z,x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \cdot \frac{x^{\nu/2-1} e^{-x/2}}{\Gamma\left(\frac{\nu}{2}\right) 2^{\nu/2}} \qquad -\infty < z < \infty, x \geq 0$$

To find the joint density of U and V, we need the *determinant* of the **Jacobian**:

$$|\mathcal{J}| = \begin{vmatrix} -\frac{1}{2^\nu}\left(\frac{x}{\nu}\right)^{-3/2} z & \left(\frac{x}{\nu}\right)^{-1/2} \\ \frac{1}{2}\left(\frac{x}{\nu}\right)^{-1/2}\left(\frac{1}{\nu}\right) & 0 \end{vmatrix} = -\frac{1}{2}x$$

So $|\mathcal{J}^{-1}| = 2x$. If we solve U in terms of Z and X, we get

$$\begin{cases} Z = UV \\ X = \nu V^2 \end{cases}$$

Thus, we can plug in these Z and X values into the joint density of Z and X as written above and multiply by the inverse of the Jacobian to get:

$$f_{U,V}(u,v) = (2\nu v^2) \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{1}{2}u^2 v^2}}{\Gamma\left(\frac{\nu}{2}\right) 2^{\nu/2}} (\nu v^2)^{\frac{\nu}{2}-1} e^{-\nu v^2/2}$$

Then, to get the *pdf* for U, we simply integrate out V over its domain.

$$f_U(u) = \frac{2}{\sqrt{2\pi}} \frac{\nu^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right) 2^{\nu/2}} \int_0^\infty v^\nu e^{-\frac{1}{2}v^2(u^2+\nu)} dv$$

To solve, let $w = v^2$ and do a bunch of algebra to get the t-distribution.

1.3 Section 4.4 - Hierarchical and Mixture Models

Example:

A particle counter is imperfect. It detects each particle independently, with probability p. The number of incoming particles in a minute has a *Poisson* distribution with parameter λ. What is the expected number of counted particles?

Let X = the number of particles detected, and $N \sim$ *Poisson*(λ) be the number of incoming particles. The, by the *law of total probability*, we have:

$$
\begin{aligned}
f(x) &= P(X=x) \\
&= \sum_{\text{all } n} P(X=x|N=n)P(N=n) \\
&= \sum_{n=x}^{\infty} \binom{n}{x} p^x (1-p)^{n-x} \frac{e^{\lambda}\lambda^n}{n!} \\
&= \sum_{n=x}^{\infty} \frac{p^x(1-p)^{n-x}e^{\lambda}\lambda^n}{x!(n-x)!} \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{n=x}^{\infty} \frac{(1-p)^{n-x}}{(n-x)!}\lambda^{n-x} \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{n=0}^{\infty} \frac{(1-p)^{n}}{n!}\lambda^{n} = \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{n=0}^{\infty} \frac{[\lambda(1-p)]^n}{n!} \\
&= \frac{(\lambda p)^x e^{-\lambda}}{x!} e^{\lambda(1-p)} = \frac{(\lambda p)^x}{x!} e^{-\lambda p} \\
X &\sim Poisson(\lambda p)
\end{aligned}
$$

So, we have: $E[X] = \lambda p$

1.3.1 Theorem: (Law of Total Expectation)

Let X and Y be two random variables. Then:

$$E[X] = E[E[X|Y]]$$

Proof:

$$
\begin{aligned}
E_Y[E(X|Y)] &= \int_{-\infty}^{\infty} E[X|Y=y]f_Y(y)dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx f_Y(y)dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dydx \\
&= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx \\
&= \int_{-\infty}^{\infty} x f_X(x)dx = E[X]
\end{aligned}
$$

Examples:

1. In the previous example, we had:

$$
\begin{aligned}
X|N = n &\sim Binomial(n, p) \\
N &\sim Poisson(\lambda) \\
E[X] = E_N[E(X|N)] &= E[Np] \\
&= pE[N] = p\lambda
\end{aligned}
$$

2. A miner is trapped in a mine. There are three possible exits.

   (a) Exit 1 leads to safety in 3 hours

   (b) Exit 2 leads back to the mine in 5 hours

   (c) Exit 3 leads back to the mine in 7 hours

   Assuming that the miner selects an exit at random each time, what is the expected length of time to reach safety? Our two random variables are:

   $$
   \begin{aligned}
   Y &= \text{ selected exit 1,2,3} \\
   X &= \text{ length of time to safety}
   \end{aligned}
   $$

   We are then interested in $E[X]$.

   $$
   \begin{aligned}
   E[X] &= E_Y[E(X|Y)] \\
   &= E[X|Y=1]\cdot\frac{1}{3} + E[X|Y=2]\cdot\frac{1}{3} + E[X|Y=3]\cdot\frac{1}{3} \\
   &= \frac{1}{3}\left(3 + [5 + E[X]] + [7 + E[X]]\right) \\
   E[X] &= 15
   \end{aligned}
   $$

Definition:

A random variable X is said to have a **mixture distribution** if values of X can be derived from more than one underlying *random variable.*

1.3.2 Generalization of the Binomial-Poisson Mixture

One example of a mixture model is the example before where $X \sim Poisson(\lambda p)$. It can be derived from

$$
\begin{aligned}
X|Y &\sim Binomial(n, p) \\
Y &\sim Poisson(\lambda)
\end{aligned}
$$

But let's generalize this. Consider a case where:

$$
\begin{aligned}
X|N &\sim Binomial(n, p) \\
N &\sim Poisson(\Lambda) \\
\Lambda &\sim exponential(\beta)
\end{aligned}
$$

Now, when wanting to find expected count, we can use *law of total expectation*:

$$\begin{aligned}
E[X] &= E_N[E(X|N)] \\
&= E_N[Np] = pE[N] \\
&= pE_\Lambda[E(N|\Lambda)] \\
&= pE[\Lambda] = p\beta
\end{aligned}$$