Lecture 20

More on Expectation

Review Exercises

1. What is the expected number of rolls to get 3 immediately followed by 4 when repeatedly rolling a fair 6-sided die?

2. If \(X, Y\) are random variables with joint PDF \(f_{X,Y}\), guess the formula for \(f_W\) where \(W = X - Y\).

3. Suppose \(X_1, \ldots, X_n\) form a random sample from a \(\text{Unif}(0, 1)\) distribution. Let \(U = \max_i (X_i)\) and let \(V = \min_i (X_i)\).

   (a) What is the probability that \(U \leq u\)?
   (b) What is the probability that \(V \geq v\)?
   (c) What is the probability that \(U \leq u\) and \(V \geq v\)?
   (d) What is the probability that \(U \leq u\) and \(V \leq v\)?
   (e) What is the joint PDF \(f_{U,V}\)?

Solutions

1. Let \(X\) denote the number of rolls to get 3 then 4, and \(Y\) the number of remaining rolls to get a 3 then four assuming your previous roll was 3. Then we have

\[
E[X] = 1 + \frac{5}{6} E[X] + \frac{1}{6} E[Y] \\
E[Y] = 1 + \frac{4}{6} E[X] + \frac{1}{6} E[Y].
\]

Solving this system gives

\[
E[X] = 1 + \frac{5}{6} E[X] + \frac{1}{6} \left( \frac{6}{5} + \frac{4}{5} E[X] \right) = \frac{6}{5} + \frac{29}{30} E[X] \implies E[X] = 36.
\]

A similar calculation shows the expected number of rolls to get two 3’s in a row is 42.

2. \(f_W(w) = \int_{-\infty}^{\infty} f_{X,Y}(t, t - w) \, dt\).

3. (a) \(P(X_1 \leq u, \ldots, X_n \leq u) = \prod_{i=1}^{n} P(X_i \leq u) = u^n\) for \(u \in (0, 1)\).
(b) \( P(X_1 \geq v, \ldots, X_n \geq v) = \prod_{i=1}^{n} P(X_i \geq v) = (1 - v)^n \) for \( v \in (0, 1) \).

(c) \( P(v \leq X_1 \leq u, \ldots, v \leq X_n \leq u) = \prod_{i=1}^{n} P(v \leq X_i \leq u) = (u - v)^n \) for \( 0 < v < u < 1 \).

(d) \( P(U \leq u, V \leq v) = P(U \leq u) - P(U \leq u, V \geq v) = u^n - (u - v)^n \) for \( 0 < v < u < 1 \).

(e) Taking partial derivatives of the joint CDF above gives

\[
\frac{\partial^2}{\partial u \partial v} F_{U,V}(u,v) = n(n-1)(u-v)^n^{-2}.
\]

We will note something now that will come up again later. Fix any \( \epsilon > 0 \). Part a) shows that \( P(|U - 1| > \epsilon) \to 0 \) as \( n \to \infty \). We say “\( U \) converges to 1 in probability”.

### Covariance and Correlation

**Definition 1 (Correlation).** If \( X, Y \) are random variables the correlation is defined by

\[
\rho(X,Y) = \frac{\text{Cov}(X - \mu_X, Y - \mu_Y)}{\sigma_X \sigma_Y} = E\left[ \frac{X - \mu_X}{\sigma_X} \cdot \frac{Y - \mu_Y}{\sigma_Y} \right].
\]

If we scale \( X \) then both the covariance and the standard deviation are scaled, so the correlation is unchanged (i.e., it is scale-invariant). In fact, we will prove the following fact:

**Theorem 2.** For any random variables \( X, Y \) with \( \sigma_X, \sigma_Y > 0 \),

\[-1 \leq \rho(X,Y) \leq 1.\]

If \( \rho(X,Y) = 1 \) then \( Y = aX + b \) with \( a > 0 \). If \( \rho(X,Y) = -1 \) then \( Y = aX + b \) with \( a < 0 \).

**Proof.** Define \( U, V \) by

\[ U = \frac{X - \mu_X}{\sigma_X} \quad \text{and} \quad V = \frac{Y - \mu_Y}{\sigma_Y}. \]

Then \( \rho(U,V) = \rho(X,Y) \) (see exercises).

\[
E[(U + V)^2] = E[U^2] + E[V^2] + 2E[UV] = \text{Var}(U) + \text{Var}(V) + 2\rho(U,V) = 2 + 2\rho(U,V)
\]

and

\[
\]

This shows

\[ 2 + 2\rho(U,V) \geq 0 \quad \text{and} \quad 2 - 2\rho(U,V) \geq 0. \]

Together these imply \(-1 \leq \rho(U,V) \leq 1\). The only way either side can be an equality is if \( E[(U + V)^2] = 0 \) or \( E[(U - V)^2] = 0 \). Note that (the following only holds with probability 1, but we will ignore this detail)

\[
E[(U + V)^2] = 0 \implies \frac{X - \mu_X}{\sigma_X} = -\frac{Y - \mu_Y}{\sigma_Y}.
\]
and
\[ E[(U - V)^2] = 0 \implies \frac{X - \mu_X}{\sigma_X} = \frac{Y - \mu_Y}{\sigma_Y} \]
giving the result.

The above is equivalent to the statement
\[ E[UV]^2 \leq E[U^2]E[V^2] \]
which is a version of the Cauchy-Schwarz inequality.

We say \( X, Y \) are uncorrelated if \( \text{Cov}(X, Y) = 0 \) or equivalently that \( \rho(X, Y) = 0 \). We proved earlier that independent random variables are uncorrelated. We will see in the exercises that the converse is false.

**Covariance and Correlation Exercises**

1. Show that \( \text{Cov} \) is a symmetric bilinear form. That is:
   (a) \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \),
   (b) \( \text{Cov}(aX, Y) = a \text{Cov}(X, Y) \),
   (c) \( \text{Cov}(X + Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y) \).
   Also show \( \text{Cov}(X + b, Y) = \text{Cov}(X, Y) \).

2. Prove that
\[ \text{Cov} \left( \frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y} \right) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = E \left[ \frac{X - \mu_X}{\sigma_X} \cdot \frac{Y - \mu_Y}{\sigma_Y} \right] . \]

3. Show that if \( U, V \) are standardized versions of \( X, Y \), respectively, then \( \rho(X, Y) = \rho(U, V) \).

4. Show that \( Y = aX + b \) with \( a \neq 0 \) implies \( |\rho(X, Y)| = 1 \).

5. Let \( X \sim \mathcal{N}(0, 1) \) and let \( Y = X^2 \).
   (a) Are \( X, Y \) independent?
   (b) Are \( X, Y \) uncorrelated?
Solutions

1. (a) Just swap $X,Y$ in definition.
   (b) Note
   \[ E[(aX)Y] - E[aXE[Y]] = a (E[XY] - E[X]E[Y]). \]
   (c) Using linearity of expectation we have
   Finally, we have

2. Since standardized random variables have zero mean, we see the last term equals the first term. To see the last equals the middle, simply factor out $\frac{1}{\sigma_X\sigma_Y}$.

3. Standardized random variables have zero mean, and standard deviation 1. Thus standardizing standardized random variables does nothing. This shows
   \[ \rho(X,Y) = \text{Cov}(U, V) = \text{Cov}\left(\frac{U - \mu_U}{\sigma_U}, \frac{V - \mu_V}{\sigma_V}\right) = \rho(U, V). \]

4. Note that
   \[ \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\text{Cov}(X, aX + b)}{|a|\sigma_X^2} = \frac{a}{|a|}, \]
   which is 1 or $-1$ depending on the sign of $a$.

5. (a) No. Knowing $X$ determines $Y$ exactly.
   (b) Yes. $E[X] = 0$ and $E[XY] = E[X^3] = 0$ since
   \[ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^3 e^{-\frac{1}{2}x^2} \, dx \]
   has an odd integrand. Thus
   \[ E[XY] - E[X]E[Y] = 0 - 0 = 0. \]

**Moment Generating Functions**

Earlier we defined the following function:

**Definition 3** (Moment Generating Function). For a random variable $X$, define the moment generating function (MGF) $M_X(t)$ by

\[ M_X(t) = E[e^{tX}]. \]

We say the moment generating function exists if the expectation $E[e^{tX}]$ exists for all $t$ in some open interval $(a,b)$ containing 0.
As we will see in the exercises, moment generating functions have many nice properties. The following is beyond the scope of this class:

**Theorem 4.** If random variables $X, Y$ have the same moment generating function in some interval $(a, b)$ containing 0, then they have the same distribution.

The above theorem essentially says that if the MGF exists, then we “un-MGF” to get the distribution back (i.e., the inverse Laplace transform exists). The reason it is called a moment generating function is the following theorem whose proof is also beyond the scope of the class:

**Theorem 5.** If $X$ has an MGF $M_X$ (i.e., the MGF of $X$ exists) then

$$M_X(t) = E[e^{tX}] = E \left[ \sum_{k=0}^{\infty} \frac{t^k X^k}{k!} \right] = \sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k].$$

Furthermore, differentiating $n$ times and evaluating at 0 gives

$$M_X^{(n)}(0) = E[X^n].$$

**Moment Generating Function Exercises**

1. Compute the MGF of $X$ for the following distributions:
   
   (a) $X \sim \text{Ber}(p)$,
   (b) $X \sim \text{Bin}(n, p)$,
   (c) $X \sim \mathcal{N}(0, 1)$.

2. Let $X$ have MGF $M_X(t)$ and $Y$ have MGF $M_Y(t)$.
   
   (a) Give the MGF of $aX + b$ where $a \neq 0$.
   (b) Give the MGF of $X + Y$ assuming $X, Y$ are independent.

3. Use MGFs to prove that if $X_1, \ldots, X_n \sim \text{Ber}(p)$ are a random sample then $X_1 + \cdots + X_n \sim \text{Bin}(n, p)$.

4. (a) Let $X \sim \mathcal{N}(\mu, \sigma^2)$. What is $M_X(t)$?
   (b) Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ be independent. What is $M_{X+Y}(t)$?

**Solutions**

1. (a) 

$$E[e^{tX}] = pe^t + (1 - p).$$
(b) 
\[ E[e^{tX}] = \sum_{k=0}^{n} e^{k t} \binom{n}{k} p^k (1 - p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1 - p)^{n-k} = (pe^t + (1 - p))^n. \]

(c) 
\[
E[e^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{x^2}{2}} \, dx \\
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}((x-t)^2 + \frac{t^2}{2})} \, dx \\
= e^{\frac{t^2}{2}}.
\]

2. (a) 
\[ M_{aX+b}(t) = E[e^{t(aX+b)}] = e^{tb} E[e^{atX}] = e^{tb} M_X(at). \]

(b) 
\[ M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} e^{tY}] = E[e^{tX}] E[e^{tY}] = M_X(t) M_Y(t). \]

More generally we have the following:

**Theorem 6.** If \( X_1, \ldots, X_n \) are independent random variables each with MGF \( M_{X_i}(t) \) and \( Y = X_1 + \cdots + X_n \) then 
\[ M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t). \]

3. Let \( X \sim \text{Ber}(p) \) and \( Y \sim \text{Bin}(n, p) \). We showed \( M_X(t)^n = M_Y(t) \). Since \( M_X(t)^n \) is the moment generating function of \( X_1 + \cdots + X_n \), the result follows.

4. (a) If \( Z \sim \mathcal{N}(0, 1) \) then \( X = \sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2) \) so 
\[ M_X(t) = e^{t \mu} e^{-\frac{\sigma^2 t^2}{2}} = e^{\mu t + \frac{\sigma^2 t^2}{2}}. \]

(b) We have 
\[ M_{X+Y}(t) = e^{t \mu_1 + \sigma_1^2 t^2/2} e^{t \mu_2 + \sigma_2^2 t^2/2} = e^{t(\mu_1 + \mu_2) + (\sigma_1^2 + \sigma_2^2) t^2/2}. \]

This shows the sum has a \( \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \) distribution.