CSE 312 24SP

Section 8: Tail Bounds, Joint Distributions, and the Law of Total Expectation

Slides made by Harrison Bay

Review

	Discrete	Continuous
Joint PMF/PDF	$P_{X,Y}(x,y) = P(X = x, Y = y)$	$f_{X,Y}(x,y) \neq P(X = x, Y = y)$
Joint support $\Omega_{X,Y}$	$\{(x,y) \in \Omega_X \times \Omega_Y : p_{X,Y}(x,y) > 0\}$	$\{(x,y) \in \Omega_X \times \Omega_Y : f_{X,Y}(x,y) > 0\}$
Joint CDF	$F_{X,Y}(x,y) = \sum_{t \le x, s \le y} p_{X,Y}(t,s)$	$F_{X,Y}(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(t,s)\,ds\,dt$
Normalization	$\sum_{x,y} p_{X,Y}(x,y) = 1$	$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$
Marginal PMF/PDF	$p_X(x) = \sum_{y} p_{(X,Y)}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$
Expectation	$E[g(X,Y)] = \sum_{x,y} g(x,y) p_{X,Y}(x,y)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy$
Requirements for Indep.	$\forall x \forall y\ p_{X,Y}(x,y) = p_X(x) p_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$	$\forall x \forall y\ f_{X,Y}(x,y) = f_X(x) f_Y(y)$ $\Omega_{X,Y} = \Omega_X \times \Omega_Y$

Review (cont.)

- **Law of Total Probability (r.v. version):** If X is a discrete random variable, then

$$\mathrm{P(A)} = \sum_{x \in \Omega_X} P(A|X = x) p_X(x)$$

(X is discrete.)

- **Law of Total Expectation (event version):** Let X be a discrete random variable, and let events $A_1, \ldots, A_n$ partition the sample space. Then,

$$\mathrm{E[X]} = \sum_{i=1}^{n} E[X|A_i] P(A_i)$$

- **Law of Total Expectation (r.v. version):** Suppose X and Y are random variables. Then,

$$E[X] = \sum_{y} E[X|Y = y] p_Y(y)$$

(Y is discrete.)

Review (cont.)

- **Conditional Distributions (discrete case):** $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$. The continuous version is the same, but with a pdf instead of a pmf.
- **Conditional Expectation (discrete case):** $E[X|Y = y] = \sum_x x p_{X|Y}(x|y)$. The continuous version is the same, but with a pdf instead of a pmf (and integral instead of sum).
  - ***Linearity of expectation still applies:*** $E[X + Y|A] = E[X|A] + E[Y|A]$
- **Covariance:** How "intertwined" are X and Y?
  - $Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$
    - Recall that $Var(X) = E[(X - E[X])(X - E[X])]$
- **We have not covered the continuous versions of the Law of Total Probability and Law of Total Expectation**, but they should be intuitive given you understand the discrete versions: again, instead of pmfs, we have pdfs, and instead of sums, we have integrals.

Review (cont.)

- **Markov's Inequality**: Let X be a nonnegative random variable and $\alpha > 0$. Then

$$\mathrm{P}(\mathrm{X} \geq \alpha) \leq \frac{E[X]}{\alpha}$$

- **Chebyshev's Inequality:** Let Y be a random variable with $E[Y] = \mu$ and $Var(Y) = \sigma^2$. Then, for any $\alpha > 0$,

$$P(|Y - \mu| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$$

- **(Multiplicative) Chernoff Bound:** Suppose $X \sim Bin(n, p)$ and $\mu = np$. Then for any $0 < \delta < 1$,

$$\mathrm{P}(\mathrm{X} \geq (1 + \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{3}}$$

$$P(X \leq (1 - \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}}$$

- **Extra tidbit**: Other useful inequalities in probability include the **union bound** and **Hoeffding's Inequality** (in case you wanted to learn more). There are tons of them.

Task 1: Kahoot!

See Task 1 in your section handout.

Task 3: Exponential Tail Bounds

- Let $X \sim Exp(\lambda)$ and $k > \frac{1}{\lambda}$.
  - (a) Use Markov's inequality to bound $P(X \geq k)$.
  - (b) Use Markov's inequality to bound $P(X < k)$.
  - (c) Use Chebyshev's inequality to bound $P(X \geq k)$.
  - (d) What is the exact formula for $P(X \geq k)$?
  - (e) For $\lambda k \geq 3$, how do the bounds given in (a), (b), and (c) compare?

Task 3 Solution

- (a) Markov's inequality is direct: $P(X \geq k) \leq \frac{E[X]}{k} = \frac{1}{\lambda k}$

Task 3 Solution

- (a) Markov's inequality is direct: $P(X \geq k) \leq \frac{1}{\lambda k}$
- (b) We can multiply on both sides (flipping the sign) to get $-P(X \geq k) \geq \frac{-1}{\lambda k}$. Adding 1 on both sides, we get $P(X < k) = 1 - P(X \geq k) \geq 1 - \frac{1}{\lambda k}$

Task 3 Solution

- (a) Markov's inequality is direct: $P(X \geq k) \leq \frac{1}{\lambda k}$
- (b) We can multiply on both sides (flipping the sign) to get $-P(X \geq k) \geq \frac{-1}{\lambda k}$. Adding 1 on both sides, we get $P(X < k) = 1 - P(X \geq k) \geq 1 - \frac{1}{\lambda k}$
- (c) Chebyshev's requires some manipulation:

$$P(X \geq k) = P\left(X - \frac{1}{\lambda} \geq k - \frac{1}{\lambda}\right) \leq P\left(\left|X - \frac{1}{\lambda}\right| \geq k - \frac{1}{\lambda}\right) \leq \frac{1}{\lambda^2\left(k - \frac{1}{\lambda}\right)^2} = \frac{1}{(\lambda k - 1)^2}$$

Task 3 Solution

- (a) Markov's inequality is direct: $P(X \geq k) \leq \frac{1}{\lambda k}$
- (b) We can multiply on both sides (flipping the sign) to get $-P(X \geq k) \geq \frac{-1}{\lambda k}$. Adding 1 on both sides, we get $P(X < k) = 1 - P(X \geq k) \geq 1 - \frac{1}{\lambda k}$
- (c) Chebyshev's requires some manipulation:

$$P(X \geq k) = P\left(X - \frac{1}{\lambda} \geq k - \frac{1}{\lambda}\right) \leq P\left(\left|X - \frac{1}{\lambda}\right| \geq k - \frac{1}{\lambda}\right) \leq \frac{1}{\lambda^2\left(k - \frac{1}{\lambda}\right)^2} = \frac{1}{(\lambda k - 1)^2}$$

- (d) Take the complement of the CDF: $P(X \geq k) = e^{-\lambda k}$

Task 3 Solution

- (a) Markov's inequality is direct: $P(X \geq k) \leq \frac{1}{\lambda k}$
- (b) We can multiply on both sides (flipping the sign) to get $-P(X \geq k) \geq \frac{-1}{\lambda k}$. Adding 1 on both sides, we get $P(X < k) = 1 - P(X \geq k) \geq 1 - \frac{1}{\lambda k}$
- (c) Chebyshev's requires some manipulation:

$$P(X \geq k) = P\left(X - \frac{1}{\lambda} \geq k - \frac{1}{\lambda}\right) \leq P\left(\left|X - \frac{1}{\lambda}\right| \geq k - \frac{1}{\lambda}\right) \leq \frac{1}{\lambda^2\left(k - \frac{1}{\lambda}\right)^2} = \frac{1}{(\lambda k - 1)^2}$$

- (d) Take the complement of the CDF: $P(X \geq k) = e^{-\lambda k}$
- (e) $e^{-\lambda k} < \frac{1}{(\lambda k - 1)^2} < \frac{1}{\lambda k}$, so Markov's inequality gives the worst bound.

Task 2(c): Tail Bounds

- Suppose $X \sim Bin(6, 0.4)$. Give an upper bound for $P(X \geq 4)$ using the Chernoff bound.

Task 2(c) Solution

- Suppose $X \sim Bin(6, 0.4)$. Give an upper bound for $P(X \geq 4)$ using the Chernoff bound.

First, we solve for the values of δ that will allow us to use the Chernoff bound. We want $4 = (1+\delta)E[X] = (1+\delta)2.4$. Solving for δ here gives us $\delta = 2/3$.

Now, we can directly plug into the Chernoff bound.

$$P(X \geq 4) = P\left(X \geq \left(1 + \frac{2}{3}\right)2.4\right) \leq e^{-\left(\frac{2}{3}\right)^2\left(\frac{E[X]}{3}\right)} = e^{\frac{-4\cdot(2.4)}{27}} \approx 0.7$$

Task 4: Robbie is Late!

Suppose the probability Robbie is late to lecture on a given day is at most 0.01. Do not make any independence assumptions.

(a) Use a union bound to bound the probability that Robbie is late at least once over a 30-lecture quarter.

(b) Use a union bound to bound the probability that Robbie is **never** late over a 30-lecture quarter.

(c) Use a union bound to bound the probability that Robbie is late at least once over a 120-lecture quarter.

Task 4 Solution

Suppose the probability Robbie is late to lecture on a given day is at most 0.01. Do not make any independence assumptions.

(a) Use a union bound to bound the probability that Robbie is late at least once over a 30-lecture quarter.

Let R_i be the event Robbie is late to lecture on day i for $i = 1,2, \dots, 30$. Then by the union bound:

$$P(late\ at\ least\ once) = P\left(\bigcup_{i=1}^{30} R_i\right) \le \sum_{i=1}^{30} P(R_i) \le \sum_{i=1}^{30} 0.01 = 0.3$$

where the second inequality is because $P(R_i) \le 0.1$.

Task 4 Solution

Suppose the probability Robbie is late to lecture on a given day is at most 0.01. Do not make any independence assumptions.

(b) Use a union bound to bound the probability that Robbie is **never** late over a 30-lecture quarter.

From the previous part, we found that $P(late\ at\ least\ once) \leq 0.3$. The complement of this is exactly the probability that Robbie is never late, i.e., $P(never\ late) = 1 - P(late\ at\ least\ once)$. Since we are subtracting at most 0.3,

$$P(never\ late) = 1 - P(late\ at\ least\ once) \geq 1 - 0.3 = 0.7$$

Task 10: Lemonade Stand

- Suppose I run a lemonade stand that costs me $100 a day to operate. I sell a single lemonade for $20. Every person who walks by my stand either buys a drink or doesn't (no one buys more than one). If it is raining, n_1 people walk by my stand, and each buys a drink independently with probability p_1. If it isn't raining, n_2 people walk by my stand, and each buys a drink independent with probability p_2. It rains each day with probability p_{rain} independently of every other day. Let X be my profit over the next week. In terms of $n_1, n_2, p_1, p_2, p_{rain}$, what is $E[X]$?

Task 10 Solution

- Let R be the event it rains. Let X_i be how many drinks I sell on day i for $i = 1, \dots, 7$. We are interested in $X = \sum_{i=1}^{7}(20X_i - 100)$.

Task 10 Solution

- Let R be the event it rains. Let X_i be how many drinks I sell on day i for $i = 1, \ldots, 7$. We are interested in $X = \sum_{i=1}^{7}(20X_i - 100)$.
  - We also know that $X_i|R \sim Bin(n_1, p_1) \Longrightarrow E[X_i|R] = n_1 p_1$ from the problem statement.
  - Similarly, $X_i|R^C \sim Bin(n_2, p_2) \Longrightarrow E[X_i|R^C] = n_2 p_2$

Task 10 Solution

- Let R be the event it rains. Let X_i be how many drinks I sell on day i for $i = 1, \dots, 7$. We are interested in $X = \sum_{i=1}^{7}(20X_i - 100)$.
  - We also know that $X_i|R \sim Bin(n_1, p_1) \Longrightarrow E[X_i|R] = n_1 p_1$ from the problem statement.
  - Similarly, $X_i|R^C \sim Bin(n_2, p_2) \Longrightarrow E[X_i|R^C] = n_2 p_2$

  We then apply the Law of Total Expectation:

  $\mu_{X_i} = E[X_i] = E[X_i|R]P(R) + E[X_i|R^C]P(R^C) = n_1 p_1 \cdot p_{rain} + n_2 p_2 \cdot (1 - p_{rain})$

Task 10 Solution

- Let R be the event it rains. Let X_i be how many drinks I sell on day i for $i = 1, \ldots, 7$. We are interested in $X = \sum_{i=1}^{7}(20X_i - 100)$.
  - We also know that $X_i|R \sim Bin(n_1, p_1) \Longrightarrow E[X_i|R] = n_1 p_1$ from the problem statement.
  - Similarly, $X_i|R^C \sim Bin(n_2, p_2) \Longrightarrow E[X_i|R^C] = n_2 p_2$

  We then apply the Law of Total Expectation:

  $\mu_{X_i} = E[X_i] = E[X_i|R]P(R) + E[X_i|R^C]P(R^C) = n_1 p_1 \cdot p_{rain} + n_2 p_2 \cdot (1 - p_{rain})$

  Then Linearity of Expectation:

$$E[X] = E\left[\sum_{i=1}^{7}(20X_i - 100)\right] = 20\left(\sum_{i=1}^{7} E[X_i]\right) - 700$$

$$= 140\big(n_1 p_1 \cdot p_{rain} + n_2 p_2 \cdot (1 - p_{rain})\big) - 700$$