5th problem set for Probability and Statistics — March 25/26

Variance

Recall:

- (Definition) $\text{var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$
- (Theorem) $\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
- (Standard Deviation) $\sigma(X) = \sqrt{\text{var}(X)}$

1. Assume that solving one problem takes X minutes, where $X \in \{1, 2, 3, 4, 5\}$. The duration is random (dependent on the weather), and the probability function is $p_X(1) = p_X(2) = 0.1$, $p_X(3) = p_X(4) = 0.2$, $p_X(5) = 0.4$. (From last week, we know that $\mathbb{E}(X) = 3.7$.) Find $\text{var}(X)$ and $\sigma(X)$.

2. Do the following *alternative definitions of variance* make sense? That is, do you learn anything interesting about random variable X when you find
(a) $\mathbb{E}(X - \mathbb{E}(X))$?
(b) $\mathbb{E}(|X - \mathbb{E}(X)|)$?
(c) $\mathbb{E}((X - \mathbb{E}(X))^k)$ (for $k = 3, 4, \dots$)?

3. Let $X \sim Bin(100, 0.5)$ and $Y \sim 10Bin(100, .05)$ (thus, $Y/10$ has a binomial distribution $Bin(100, .05)$). Compute $\mathbb{E}(X)$, $\text{var}(X)$, $\sigma(X)$ and the same for Y.

4. Let $X \sim Pois(\lambda)$. Show that $\mathbb{E}(X) = \lambda$ and $\text{var}(X) = \lambda$.

Joint distribution

Recall: For n discrete random variables $X_1, \dots, X_n : (\Omega, F, P) \to \mathbb{R}$ their joint probability mass function is $p_{X_1,\dots,X_n} : \mathbb{R}^n \to [0, 1]$,

$$p_{X_1,\dots,X_n}(x_1, \dots, x_n) = P(X_1 = x_1 \wedge \dots \wedge X_n = x_n)$$

Marginal distribution: For discrete rv's X, and Y with joint pmf $p_{X,Y}$, the marginal of X is

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

5. We draw two cards from the standard 52-card deck. Let X denote the number of aces drawn and Y denote the number of kings drawn. Determine the associated probability function $p_{X,Y}$ and the marginals p_X and p_Y.

6. We toss a fair coin three times. Let X be the number of heads in the first two tosses, and Y be the number of tails in the last two tosses.
(a) Determine the joint probability function $p_{X,Y}$ and also marginals p_X, p_Y.
(b) Are X and Y independent?
(c) Determine $P(X < Y)$.
(d) Determine the conditional probability function $p_{X|Y}$, i.e., the numbers $P(X = x \mid Y = y)$ for all values of x, y.

7. We say that the random variables X and Y are uncorrelated if the covariance $Cov(X, Y) = 0$, where the covariance is defined as $Cov(X, Y) := \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. Define the random variable X that takes values in $\{1, 2, 3\}$ with $P(X = i) = 1/3$ for each $i \in \{1, 2, 3\}$. Consider the random variable $Y = 1$, if $X = 2$ and $Y = 0$, otherwise. Show that X and Y are uncorrelated. Are X and Y independent random variables?

8. Let X, Y be two independent random variables having Poisson distribution with parameters λ and μ respectively.
(a) Find the distribution of the convolution $Z = X + Y$.
(b) Calculate $\mathbb{E}(Z(Z - 1))$. (Hint: Use LOTUS.)
(c) Compute $\text{var}(Z)$.
(d) If $\mu = 2\lambda$, compute $Cov(X, Z)$.

Recognizing Random Variables

9. The probability of a data breach at our server for each given day is 0.01, independently for each day. Let T be the number of days until the first data breach.

(a) What is the distribution of T?

(b) Calculate $\mathbb{E}(T)$ and $\mathrm{var}(T)$.

(c) What is the probability that the server remains secure for an entire year?

10. Each software test can either find a bug (which we count as success) or not (this we count as failure). Assume the probability of finding a bug in one test is 0.05, and a developer performs 20 independent tests. Let X be the number of bugs found.

(a) What is the distribution of X?

(b) Calculate $\mathbb{E}(X)$ and $\mathrm{var}(X)$.

(c) What is the probability of finding exactly three bugs?

More Practice Problems

11. (Jensen's inequality) Let g be a convex function and X be a random variable. Prove that $\mathbb{E}(g(X)) \geq g(\mathbb{E}(X))$. [Hint: Assume g(X) is a convex function and the linear function $f(X) = a + bX$ is tangential to $g(X)$ at the point $\mathbb{E}(X)$. Recall that the graph of a convex function lies entirely above its tangent at every point. Now, use the linearity of expectation.]

12. * Consider a permutation of $\{1, 2, \ldots, n\}$ chosen uniformly at random from all possible permutations (as before). Let the random variable X be the number of fixed points in this random permutation. Find $\mathbb{E}(X)$ and $\mathrm{var}(X)$.

13. Let's consider a group of m married couples (i.e., a total of $2m$ individuals). Suppose that after ten years, each of these $2m$ people will still be alive with probability p, independently of the others. We do not consider possibilities of divorce, etc., so the couples are immutable.

Let L be the set of people who will be alive after ten years, and A their number (i.e., $A = |L|$). Furthermore, let B be the number of couples where both partners will be alive; thus, A, B are random variables satisfying $0 \leq A \leq 2m$ and $0 \leq B \leq m$. For each $a = 0, \ldots, 2m$, we want to compute $\mathbb{E}(B \mid A = a)$.

(a) Let's consider a specific individual. What is the probability that they will be alive after ten years, given that $A = a$? In other words, if that person is x, what is $P(x \in L \mid A = a)$?

(b) Let's consider a specific married couple. What is the probability that both partners will be alive, given that $A = a$?

(c) Express B as the sum of m suitable indicator random variables.

(d) The linearity of expected value also holds for conditional expected value, i.e., $\mathbb{E}(\sum_{i=1}^{m} X_i \mid J) = \sum_{i=1}^{m} \mathbb{E}(X_i \mid J)$, for any event J and random variables $X_1, \ldots, X_m$. Utilize this to compute $\mathbb{E}(B \mid A = a)$.

(e) What is the distribution of random variable A? (Either name it or write the probability function, i.e., determine $P(A = a)$.)

(f) For a chosen a-element set of people M, what is the probability that it exactly corresponds to the set of survivors? In other words, what is $P(L = M)$? And what about $P(L = M \mid A = a)$?