Statistics 100A
Homework 4 Solutions

Ryan Rosario

Problem 1

For a discrete random variable X,

Note that all of the problems below as you to *prove* the statement. We are proving the properties of expectation and variance, thus you cannot just use them right off the bat.

(1) *Prove* $E(aX+b) = aE(X)+b$.

Proof. By definition,

$$E(X) = \sum_x xp(x)$$

where f is a discrete probability distribution. Thus,

$$\begin{aligned}
E(aX+b) &= \sum_x (ax+b)p(x) \\
&= \sum_x axp(x) + \sum_x bp(x) \\
&= a\sum_x xp(x) + b\sum_x p(x) \\
& \text{Note that } E(X) = \textstyle\sum_x xp(x) \text{ and } \sum_x p(x) = 1. \\
&= aE(X) + b
\end{aligned}$$

This proves linearity of expectation with respect to one variable.

□

(2) *Prove* $Var(aX+b) = a^2 Var(X)$.

First, prove a lemma: $\text{Var}(X) = E(X^2) - E(X)^2$.

Proof. By definition, $\text{Var}(X) = E\left[(X-\mu)^2\right]$.

$$\begin{aligned}
\text{Var}(X) &= E\left[(X-\mu)^2\right] \\
&= E\left[X^2 + \mu^2 - 2\mu X\right] \\
& \text{Recall that } \mu = E(X). \\
&= E\left[X^2 - \mu^2\right] \\
& \text{By part 1, linearity:} \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - E(X)^2
\end{aligned}$$

$\square$

Then we can work with $\text{Var}(aX + b) = a^2\text{Var}(X)$.

Proof. We can use the result from part 1 and the lemma to solve this problem.

$$\begin{aligned}
\text{Var}(aX+b) &= E\left[\{(aX+b) - (a\mu - b)\}^2\right] \\
&= E\left[a^2(X-\mu)^2\right] \\
&= a^2\left(E\left[X^2+\mu^2-2\mu X\right]\right) \\
&= a^2\left(E(X^2)-\mu^2\right) \\
&= a^2\left(E(X^2)-E(X)^2\right) \\
&= a^2\text{Var}(X)
\end{aligned}$$

$\square$

Or, we can start with basic principles. Let $Z = aX + b$.

Proof.

$$\begin{aligned}
\text{Var}(Z) &= E(Z^2) - E(Z)^2 \\
&= \left(\sum_x (aX+b)^2 f(x)\right) - (aX+b)^2 \\
&= \left(\sum_x \left[(aX)^2 + b^2 + 2abX\right] f(x)\right) - \left[(aE(X))^2 + b^2 + 2abE(X)\right] \\
&= \left(a^2\sum_x X^2 f(x) + b^2\sum_x f(x) + 2ab\sum_x xf(x)\right) - \left[a^2E(X)^2 + b^2 + 2abE(X)\right] \\
&= a^2E(X^2) + b^2 + 2abE(X) - a^2E(X)^2 - b^2 - 2abE(X) \\
&= a^2E(X^2) - a^2E(X)^2 \\
&= a^2\left(E(X^2) - E(X)^2\right) \\
&= a^2\text{Var}(X)
\end{aligned}$$

$\square$

(3) *Let* $\mu = E(X)$ *and* $\sigma^2 = Var(X)$. *Let* $Z = \frac{X-\mu}{\sigma}$, *calculate* $E(Z)$ *and* $Var(Z)$.

We want to find the expected value and variance of Z where $Z = \frac{X-\mu}{\sigma}$. This is called a z score, and is the primary metric associated with the normal distribution. The z scores measures how far from the mean, in standard deviations, a particular value x is.

$$\begin{aligned}
E(Z) &= E\left(\frac{X-\mu}{\sigma}\right) \\
&\quad \text{By linearity (part 1), we get:} \\
&= E\left(\frac{X}{\sigma}\right) - E\left(\frac{\mu}{\sigma}\right) \\
&\quad \text{But since } \mu \text{ and } \sigma \text{ are constants, they can be pulled out of the expectation by linearity (part 1).} \\
&= \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} \\
&\quad \text{But since } E(X) = \mu \text{ we get} \\
&= \frac{\mu}{\sigma} - \frac{\mu}{\sigma} \\
&= 0
\end{aligned}$$

By definition, $\text{Var}(X) = E\left[(X-\mu)^2\right]$.

$$\begin{aligned}
\text{Var}(Z) &= E\left[(Z-\mu_Z)^2\right] \\
&\quad \text{Where } \mu_Z = E(Z) = 0 \text{ from the previous part.} \\
&= E\left[Z^2\right] \\
&= E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] \\
&= E\left[\frac{X^2+\mu^2-2\mu X}{\sigma^2}\right] \\
&= \frac{1}{\sigma^2}\left[E(X^2)+\mu^2-2\mu E(X)\right] \\
&\quad \text{But } E(X) = \mu \text{ so} \\
&= \frac{1}{\sigma^2}\left[E(X^2)-\mu^2\right] \\
&\quad \text{From the lemma proven in part 2,} \\
&= \frac{1}{\sigma^2}\text{var}(X) \\
&\quad \text{and since } \sigma^2 = \text{Var}(X), \\
&= \frac{\sigma^2}{\sigma^2} \\
&= 1
\end{aligned}$$

We have proven that $Z \sim N(0,1)$. $N(0,1)$ is called the standard normal distribution.

(4) *Prove* $Var(X) = E(X^2) - E(X)^2$.

This was the lemma that we proved in part 2.

Problem 2

Suppose in the popultion of voters, the proportion of those who would vote for a certain candidate is 20%.

(1) *If we randomly sample 5 people from the population of voters, what is the probability that we get at least 2 supporters of this candidate?*

Each person's support is independent. The probability that a person supports this candidate is 0.2. Let X be the number of people that support this candidate, $n = 5$ and $p = 0.2$.

$$X \sim \text{Binomial}(n, p)$$

Then,

$$\begin{aligned}
P(X \geq 2) &= 1 - P(X \leq 1) \\
&= 1 - [P(X = 0) + P(X = 1)] \\
&= 1 - \left[\begin{pmatrix} 5 \\ 0 \end{pmatrix} (0.2)^0 (0.8)^5 + \begin{pmatrix} 5 \\ 1 \end{pmatrix} (0.2)^1 (0.8)^4 \right] \\
&= 1 - \left((0.8)^5 + 5(0.2)^1 (0.8)^4 \right) \\
&= 0.26272
\end{aligned}$$

Of course, computing $P(X \geq 2)$ brute force is also acceptable.

(2) *If we randomly sample 100 people from the population of voters. Let X be the number of people among these 100 people who support the candidate. What is the probability that $X > 28$? You only need to write down this probability. You do not need to calculate it.*

Now we have $n = 100$ and $p = 0.2$. So,

$$P(X > 28) = \sum_{k=28}^{100} \begin{pmatrix} 5 \\ k \end{pmatrix} 0.2^k 0.8^{100-k}$$

Aside: We can calculate this on a computer in software such as R:

```
n <- 100
p <- 0.2
k <- seq(29,100)
sum(choose(n,k)*(p**k)*((1-p)**(n - k)))
[1] 0.02002021
```

Thus, the exact answer is $P(X > 28) \approx 0.02$.

Aside: There is a way to compute this directly. Since n is large, and p is not too close to 0 or 1 (or, from Stats 10, since $np > 10$ and $n(1-p) > 10$) we can use the normal distribution to approximate the normal distribution where $\mu = np = 100 \cdot 0.2 = 20$ and $\sigma = \sqrt{np(1-p)} = \sqrt{16} = 4$.

$$
\begin{aligned}
P(X > 28) &= P\left(Z > \frac{28-20}{4}\right) \\
&= P\left(Z > \frac{8}{4}\right) \\
&= P\left(Z > 2\right) \\
&= 1 - P\left(Z < 2\right) \\
&\quad \text{Which can be found in a } z \text{ table or by using R:} \\
&\approx 0.0228
\end{aligned}
$$

Note that the normal approximation is fairly close, but with an error of 10^{-4} which is OK but not great.

(3) *What is $E(X)$ and Var(X)? What is the standard deviation of X?*

There is an ambiguity in this problem. It is not clear which n to use from the problem.

In any case, $E(X) = np$ and $\mathrm{Var}(X) = np(1-p)$ and $\sigma_x = \sqrt{np(1-p)}$.

	$n = 5$	$n = 100$
$E(X)$	1	20
$\mathrm{Var}(X)$	0.8	16
σ_x	$\sqrt{0.8} = 0.984$	4

Problem 3

For $T \sim$ Geometric(p), calculate $E(T)$. If $p = .2$, then what is the probability that $T > 5$?

We need to derive $E(T)$ from basic principles. Recall that there are two parameterizations for the geometric distributions.

(1) If t is the number of failures before the first success, then $P(T = t) = (1-p)^t p$.

(2) If the tth attempt is the first success, then $P(T = t) = (1-p)^{t-1} p$.

By definition,

$$E(T) = \sum_t tP(T = t)$$

Using the second parameterization,

$$\begin{aligned}
E(T) &= \sum_{t=1}^{\infty} tP(T = t) \\
&= \sum_{t=1}^{\infty} t(1-p)^{t-1}p \\
&= p\sum_{t=1}^{\infty} t(1-p)^{t-1} \\
&\quad \text{Trick here:} \\
&= p\left[\sum_{t=1}^{\infty}(1-p)^{t-1} + \sum_{t=2}^{\infty}(1-p)^{t-1} + \ldots\right] \\
&= p\left[\{1 + (1-p) + (1-p)^2 + \ldots\} + \{(1-p) + (1-p)^2 + \ldots\} + \ldots\right] \\
&\quad \text{Each term is a convergent geometric series: } \Sigma = \tfrac{a_0}{1-\frac{a_{n+1}}{a_n}} \\
&= p\left[\frac{1}{1-(1-p)} + \frac{1-p}{1-(1-p)} + \ldots\right] \\
&= p\left[\frac{1}{p} + \frac{1-p}{p} + \ldots\right] \\
&= 1 + (1-p) + (1-p)^2 + \ldots \\
&\quad \text{which is a geometric series itself, so} \\
&= \frac{1}{1-(1-p)} \\
&= \frac{1}{p}
\end{aligned}$$

The proof is very similar for the other parameterization, and there are other ways to do this. One common way is by using derivatives.

Now to compute $P(T > 5)$. If we assume that t is the number of failures until the first success, then

$$P(T > 5) = (1-p)^5 p + (1-p)^6 p + \ldots = \frac{(1-p)^5 p}{1-(1-p)} = (1-p)^5$$

Thus,

$$P(T > 5) = (0.8)^5 \approx 0.328$$

In general, assuming t is the number of failures until the first success, then $P(T > t) = (1-p)^t$.

Problem 4

Suppose we divide the time axis into small periods $(0, \Delta t), (\Delta t, 2\Delta t), \ldots$ *Within each period, we flip a coin independently. The probability of getting a head is* $\lambda \Delta t$.

(1) *Let X be the number of heads within the interval $[0, t]$. Calculate the limit of $P(X = k)$ as $\Delta t \to 0$, for $k = 0, 1, 2, \ldots$. Also calculate $E(X)$.*

Within each subinterval, we flip a coin. The probability of getting a head (a success) is $\lambda \Delta t$ and X is the number of heads that we observe in the interval $[0, t]$. Note that each subinterval forms a Bernoulli trial and the sequence of trials is independent. Start with the binomial distribution.

$$P(X = k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}$$

where the probability of success (head) is $\lambda \Delta t$ and the number of trials (subintervals) is $\frac{t}{\Delta t}$. Thus, $n = \frac{t}{\Delta t}, p = \lambda \Delta t$.

The first trick that we need to use is the following

$$(1-p)^{n-k} \approx e^{-np}$$

Proof. If we take the log of $(1-p)^{n-k}$ we get $\log{(1-p)^{n-k}} = (n-k)\log{(1-p)}$. Note that for p very small, $p << 1$, we can approximate that $\log{(1-p)} \approx -p$. Then,

$$\begin{aligned}
\log{(1-p)^{n-k}} &= (n-k)\log{(1-p)} \\
&\approx (n-k)(-p) \\
&= -p(n-k) \\
& \text{Thus...} \\
(1-p)^{n-k} &\approx e^{-p(n-k)} \\
&\approx e^{-np} \\
&= e^{-\lambda t}
\end{aligned}$$

□

Now recall that

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k!(n-k)!}$$

Another trick we can use is that $\frac{n!}{(n-k)!} \approx n^k$.

Starting with the binomial distribution, we have that $\frac{n!}{(n-k)!} \approx n^k$ and $(1-p)^{n-k} \approx e^{-np}$, thus in the limit

$$P(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

(2) *Let T be the time until the first head. Calculate the limit of $P(T > t)$ as $\Delta t \to 0$. In both (1) and (2), let us assume that t is a multiple of λt.*

We want to find the probability that the first success occurs after the kth attempt. This is geometric. Recall that $n = \frac{t}{\Delta t}, p = \lambda t$.

$$P(T > k) = (1-p)^k = (1 - \lambda \Delta t)^{\frac{t}{\Delta t}}$$

Recall that by definition, $\left(1 + \frac{x}{n}\right)^n \to e^x$, so $(1 - \lambda \Delta t)^{\frac{t}{\Delta t}} \to e^{-\lambda t}$

(3) *Please give a real example of X and T with concrete value of λ and its unit. What is the meaning of λ.*

Answers will vary.

In general, λ is the rate parameter and specifies the (average) number of events that occur per some unit time (seconds, hours, etc.) Everything, λ, X, and T must all use this same unit of time.

X is a random variable that represents the number of events that happen during *some* unit time.

T is a random variable that represents the time it takes for the first event to occur.

Note that $X \sim$ Poisson(λ). Since T is the amount of time for the first event to occur, $T \sim$ Geometric. If T where the amount of time *between* two events, then T would be exponential.

One common example is a call center. Let's use cell phones since this is 2011. In the call center problem, we have that a particular cell tower receives on average 20 phone calls *per fifteen minutes*. The unit of time here is *fifteen minutes*. Everything in the problem must then be written in terms of *fifteen minutes*. Let X be the number of calls received during some arbitrary *fifteen minute* block of time. Then T represents the amount of time (in *fifteen minute* blocks) until the cell tower receives the first call.

Catch: If we wanted to find the probability that the cell tower receives 70 calls in an *hour*, we need to rewrite it with respect to fifteen minute blocks. This is where λt comes in. The average number of calls *per hour* is $\lambda t = 20 \cdot 4 = 80$ because there are 4 *fifteen minute* blocks in one hour. Then,

$$P(X = 70) = \frac{(80)^{70} e^{-80}}{70!} = 0.025$$