stat400 / exam02

This is a test

3.3 ~ expected value (discrete rvs).

Given some discrete random variable X that has pf $p(x)$ and set of possible values D, the **expected value** of X is the weighted average of all possible values of X:

$$E(x) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Sometimes we want to find the expected value of a function of a random variable, or $E(h(X))$ for some function h. Deriving the following formula is pretty simple:

$$E(h(X)) = \sum_{x \in D} h(x) \cdot p(x)$$

In the case of a linear function $h(X) = aX + b$, we can use a shortcut formula. This is also easily derivable by manipulating the sum in the $E(x)$ definition.

$$E(aX + b) = aE(x) + b$$

The **variance** of some rv X is a measure of the distribution's variability. It is also the square of the **standard deviation**.

$$\sigma_x^2 = V(X) = E[(x - \mu_x)^2] = E(X^2) - [E(X)]^2$$

The square root of the variance is the **standard deviation**.

$$\sigma_x = \sqrt{V(X)}$$

To find the variance of a linear function $aX + b$, use the following formula:

$$V(aX + b) = a^2 \cdot V(x)$$
$$\sigma_{aX+b} = |a|\sigma_x$$

We need the absolute value in the above: even if a is negative, the distances of possible outcomes from the mean are always positive.

3.4 ~ binomial distribution.

The **binomial distribution** is used to describe a random variable that measures n trials, X of which are successful. p is the probability of a trial being successful. The probability function is as follows:

$$b(X = x; n, p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x \in \mathbb{N} \\ 0 & \text{else} \end{cases}$$

The cdf of the binomial distribution is just:

$$P(X \leq x) = \sum_{s=0}^{x} b(s; n, p) \qquad x \in \mathbb{N}$$

The mean of the binomial distribution $E(X) = np$, and the variance $V(x) = np(1-p)$.

3.6 ~ poisson distribution.

A random variable X has a **poisson distribution** if its pmf is:

$$p(x; \mu) = \frac{e^{-\mu}\mu^x}{x!}$$

Note that while μ is the parameter of the distribution, it is also its expected value. In the poisson distribution, $E(x) = V(x) = \mu$.

Suppose we have a binomial distribution with parameters n and p. If choose $n \to \infty$ and $p \to 0$ in such a way that $np \to \mu$, then we can say that:

$$b(x; n, p) \to p(x; \mu)$$

Another important use of this distribution is the **poisson process**, the occurrence of events over time. The process has a parameter $\alpha > 0$ such that the expected number of events over time t is αt.

The below pdf demonstrates the probability that k events will occur according to the poisson process:

$$P_k(t) = \frac{e^{-\alpha t}(\alpha t)^k}{k!}$$

4.1 ~ continuous pdfs.

Continuous random variables are those that can take on an infinite number of values. For instance, the height of all American men.

If X is a continuous random variable with pdf $f(x)$, the probability that X lies between a and b can be determined by:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

For $f(x)$ to be a legitimate pdf, it must meet two conditions.

1. $f(x) \leq 0, \forall x$. No negative probabilities.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$. Probabilities of all outcomes sum up to 1.

If a continuous random variable has a **uniform distribution** on the interval $[A, B]$, it has the following pdf:

$$f(x; A, B) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

Note that the probability of a continuous random variable taking on an exact value is zero:

$$\forall c, P(X = c) = 0$$

4.2 ~ continuous cdfs and evs.

The **cumulative distribution function (cdf)** of a continuous random variable X is the probability that $X \leq x$, where x is a parameter of the function:

$$\text{cdf of } X = F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

The reason we use y inside the integral is because x is already a parameter of the function; the choice of y is arbitrary but probably convention or something.

We can use the cdf to compute various probabilities:

$$P(X > x) = 1 - F(x)$$
$$P(a \leq X \leq b) = F(b) - F(a)$$

Since $F(x)$, the cdf by definition computes the integral of $f(x)$, the pdf, we can determine $f(x)$ by taking the derivative of $F(x)$.

$$F'(x) = f(x) \text{ at every } x \text{ where } F'(x) \text{ exists}$$

The **expected value** of a continuous rv X is found with the following:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

If we want to find the expected value of some function $h(X)$, we compute:

$$E(h(X)) = \int_{-\infty}^{\infty} h(x) \cdot f(x)dx$$

To find the the variance of some continuous rv X. (Standard deviation $\sigma = \sqrt{V(x)}$.)

$$V(x) = \int_{\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx = E((x - \mu)^2)$$

Shorthand to calculate the variance:

$$V(x) = E(X^2) - (E(X))^2$$

4.3 ~ normal distribution.

A continuous rv X is said to have a **normal distribution** with parameters μ and σ, where $-\infty < \mu < \infty$ and $\sigma > 0$ if it has the following pdf:

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

The expected value is equal to the parameter μ, and the distribution's standard deviation is equal to the parameter σ.

It follows that finding $P(a \leq X \leq b)$ is dependent on integrating the above function:

$$P(a \leq X \leq B) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

The standard normal distribution is one that has $\mu = 0$ and $\sigma = 1$, denoted by Z:

$$f(z, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

We can "standardize" other normal distributions using the factor. The center of X can be shifted to zero by subtracting by μ, after which we can scale it by $\frac{1}{\sigma}$ so its standard deviation is 1.

$$Z = \frac{X - \mu}{\sigma}$$

>.< ~ practice midterm.

Question 1: Let X be the outcome of rolling a fair six-sided die. What is the expected value of X?

> *X can take any number between 1 and 6. There is a $\frac{1}{6}$ chance of each:*

> $$E(X) = \frac{1}{6}\sum_{i=1}^{6} x = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

Question 2: A chemical supply company currently has in stock 100 lb of a certain chemical, which it sells to customers in 5 lb batches. Let X be the number of batches ordered by a randomly chosen customer and suppose that X has probability (moment) function:

x	1	2	3	4
$f(x)$	0.2	0.4	0.3	0.1

Compute $E(X)$ and $V(X)$. Then, compute the expected number of pounds left after the next customer's order is shipped and the variance of the number of pounds left.

Hint: The number of pounds left is a linear function of X.

> *First, let's find the expected value:*
>
> $$E(X) = 1(0.2) + 2(0.4) + 3(0.3) + 4(0.1) = 2.3$$
>
> *Next, we find the variance:*
>
> $$V(X) = E[(X-\mu)^2] = (1-2.3)^2(0.2) + (2-2.3)^2(0.4) + (3-2.3)^2(0.3) + (4-2.3)^2(0.1) = 0.81$$
>
> *Lastly, let's find the expected number of pounds left. The number of pounds remaining after X batches have been sold can be modeled with the following function:*
>
> $$h(X=x) = 100 - 5x$$
>
> *Let's find the expected value of this function:*
>
> $$E(h(X)) = 100 - 5E(X) = 100 - 5(2.3) = 88.5$$
>
> *And finally, its variance as well:*
>
> $$V(h(X)) = 5^2 V(x) = 25(0.81) = 20.25$$

Question 3: Let f_n denote the probability (moment) function of a $\text{Binom}(n, \frac{1}{n})$-distributed random variable. Let p_λ denote the probability (moment) function of a $Poi(\lambda)$-distributed random variable. Which of the following statements is true for all $x \in \mathbb{R}$?

1. $\lim_{n\to\infty} f_n(x) = p_0(x)$
2. $\lim_{n\to\infty} f_n(x) = p_1(x)$
3. $\lim_{n\to\infty} f_n(x) = p_n(x)$
4. None of the other statements are true.

> *For our binomial distribution to approximate our poisson distribution, np must approach some λ as n approaches infinity. In our case, p is equal to $\frac{1}{n}$, so $np = n \cdot \frac{1}{n} = 1$. In other words, as $n \to \infty$, $np \to 1$, so $f_n(x)$ approaches $p_1(x)$.*
>
> ***Statement 2 is true.***

Question 4: Suppose that trees are distributed in a forest according to a two-dimensional Poisson process with parameter α, the expected number of trees per acre, equal to 80.

4a: What is the probability that in a certain quarter-acre plot, there will be at most 16 trees?

> *Let's write the pdf for our Poisson process:*
>
> $$P_k(t) = \frac{e^{-80t}(80t)^k}{k!}$$
>
> *We want to calculate the probability that k, the number of trees, is between 0 and 16, inclusive, when $t = 0.25$, number of acres.*
>
> $$P(0 \leq k \leq 16) = \sum_{k=0}^{16} \frac{e^{-80 \cdot 0.25}(80 \cdot 0.25)^k}{k!} \approx 0.2211$$

4b: If the forest covers 85000 acres, what is the expected number of trees in the forest?

> *The expected value of a poisson distribution is the parameter, αt in our case.*
>
> $$\alpha t = 80 \cdot 85000 = 6800000$$

4c: Suppose you select a point in the forest and construct a circle of radius 0.1 mile. Let X be the number of trees within that circular region. What is the probability (moment) function of X? Hint. One square mile is 640 acres.

> *If the radius of the circle is 0.1 miles, the area must be $\pi(0.1)^2 = 0.01\pi$ square miles, which is 6.4π acres. αt is then $80 \cdot 6.4\pi = 512\pi$.*
>
> *The probability function of X would then be:*
>
> $$P(X = x) = \begin{cases} \frac{e^{-512\pi}(512\pi)^x}{x!} & x \in \mathbb{N}_0 \\ 0 & \text{else} \end{cases}$$

Question 5: Let $X \sim \text{Unif}[0, 1]$. What is the standard deviation of X?

> *Let's first find $E(X)$:*
>
> $$\begin{aligned} E(X) &= \int_{-\infty}^{0} 0(x)dx + \int_{0}^{1} \frac{1}{1-0}(x)dx + \int_{1}^{\infty} 0(x)dx \\ &= \int_{0}^{1} xdx = \left[\frac{1}{2}x^2\right]_0^1 = \frac{1}{2} \end{aligned}$$
>
> *Next, let's find $E(X^2)$:*
>
> $$\begin{aligned} E(X) &= \int_{-\infty}^{0} 0(x^2)dx + \int_{0}^{1} \frac{1}{1-0}(x^2)dx + \int_{1}^{\infty} 0(x^2)dx \\ &= \int_{0}^{1} x^2dx = \left[\frac{1}{3}x^3\right]_0^1 = \frac{1}{3} \end{aligned}$$
>
> *Let's use the shorthand for variance to find $V(X)$:*
>
> $$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$
>
> *Finally, we can find the square root:*
>
> $$\sigma = \sqrt{V(X)} = \sqrt{\frac{1}{12}} = \frac{1}{2\sqrt{3}}$$

Question 6: Let X be a continuous random variable with cumulative distribution function:

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \frac{x}{4}(1 + \ln \frac{4}{x}) & \text{if } 0 < x \leq 4, \\ 1 & \text{if } x > 4. \end{cases}$$

6a: Find $P(X \leq 1)$.

> *By the definition of a cdf, we know that* $P(X \leq 1) = F(1)$.
>
> $$P(X \leq 1) = F(1) = \frac{1}{4}(1 + \ln \frac{4}{1}) = 0.5966$$

6b: Find $P(1 \leq X \leq 3)$.

> *We can rewrite* $P(1 \leq X \leq 3)$ *using our cdf:*
>
> $$P(1 \leq X \leq 3) = F(3) - F(1)$$
>
> *We found* $F(1)$ *in* ***(a)****, let's evaluate* $F(3)$*:*
>
> $$F(3) = \frac{3}{4}(1 + \ln \frac{4}{3}) = 0.9658$$
>
> *Finally,*
>
> $$P(1 \leq X \leq 3) = F(3) - F(1) = 0.9658 - 0.5966 = 0.3692$$

6c: Find the probability density function of X?

> *We can determine the pdf of* X *by taking the derivative of the cdf, for intervals where the derivative exists.*
>
> *For* $x < 0$*, we know that* $F(x) = 0$*. As a result, the pdf of* x *in this interval must be:*
>
> $$\frac{d}{dx}[0] = 0$$
>
> *For* $0 < x \leq 4$*, we know that* $F(x) = \frac{x}{4}(1 + \ln \frac{4}{x})$*. The pdf in this interval must be:*
>
> $$\begin{aligned} \frac{d}{dx}\left[\frac{x}{4}(1 + \ln \frac{4}{x})\right] &= \frac{1}{4}(1 + \ln \frac{4}{x}) + \frac{x}{4}(-\frac{4}{x^2})(\frac{x}{4}) \\ &= \frac{1}{4} + \frac{1}{4}\ln \frac{4}{x} - \frac{1}{4} \\ &= \frac{1}{4}\ln \frac{4}{x} \end{aligned}$$
>
> *For* $x > 4$*, we know that* $F(x) = 1$*. The pdf in this interval must be:*
>
> $$\frac{d}{dx}[1] = 0$$
>
> *Let's now make sure that* $F(x)$ *is differentiable at points between intervals,* $x = 0, x = 4$*.*
>
> *For* $x = 0$*:*
>
> $$0 = 0$$
> $$\frac{1}{4}\ln \frac{4}{0} = \text{undefined}$$
>
> *So we cannot determine* $f(0)$*.*

> *For $x = 4$:*
>
> $$\begin{aligned} \frac{1}{4} \ln \frac{4}{4} &= 0 \\ 0 &= 0 \end{aligned}$$
>
> *The two expressions are equivalent, so $f(4)$ exists and equals 0.*
>
> *We have found a valid pdf expression for all intervals, so altogether:*
>
> $$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{4}(\ln \frac{4}{x} + 5) & \text{if } 0 < x \leq 4 \\ 0 & \text{if } x > 4 \end{cases}$$

>!< ~ practice midterm score.

Question 1: $2/2$, Just taking expected value of a discrete random variable.

Question 2: $6/6$, Technically I didn't use the shortcut method to find the variance (there was a point associated with finding $E(X^2)$) but it was not required so I'm still giving myself full credit. All my answers were correct.

Question 3: $2/2$, Remember how Binom can approach Poi ($np \to \mu$).

Question 4: $8/9$. I didn't write the full form of the pmf. Make sure to include that any excluded intervals have a 0 probability.

Question 5: $2/2$. Know how to determine the variance of a continuous rv. I used the shortcut method and did integrals to find the relevant expected values, $E(X)$ and $E(X^2)$.

Question 6: $5/8$. Sold finding the pdf from cdf. Things to remember:

1. Find the derivative over each interval (did this).
2. Make sure the shared point between intervals has equal $F'(x)$ on both sides (didn't do this).
    1. If they are equal, then you can include it in the pdf.
    2. Otherwise, you can't; cdf is not differentiable at that point.
3. Actually take the derivative correctly (did not do this).

Final Score: $25/29 \approx 86\%$.