Lecture 16: Sept 28

Last time

- HW3 posted
- Midterm Exam 1 10/10, will have a practice exam
- Probability integral transformation
- Expectations (2.2)

Today

- Exam 1 covers up to next Monday's lecture
- Expectations (2.2)
- Moments and moment generation function

Expectation

The process of taking expectations is a linear operation, which means that the expectation of a linear function of X can be easily evaluated by noting that for any constants a and b, such that

$$E(aX + b) = aEX + b$$

Theorem Let X be a random variable and let a, b, and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,

1. $E(ag_1(X) + bg_2(X) + c) = aEg_1(X) + bEg_2(X) + c$.
2. If $g_1(x) \geqslant 0$ for all x, then $Eg_1(X) \geqslant 0$.
3. If $g_1(x) \geqslant g_2(x)$ for all x, then $Eg_1(X) \geqslant Eg_2(X)$.
4. If $a \leqslant g_1(x) \leqslant b$ for all x, then $a \leqslant Eg_1(X) \leqslant b$.

Proof:
We will give details for only the continuous case, the discrete case being similar. By definition

$$\begin{aligned} E(ag_1(X) + bg_2(X) + c) &= \int_{-\infty}^{\infty} \left[ag_1(x) + bg_2(x) + c\right] f_X(x)dx \\ &= \int_{-\infty}^{\infty} ag_1(x)f_X(x)dx + \int_{-\infty}^{\infty} bg_2(x)f_X(x)dx + \int_{-\infty}^{\infty} cf_X(x)dx \\ &= aEg_1(X) + bEg_2(X) + c \end{aligned}$$

The other three properties are proved in a similar manner (shown in class).

Example (Method of indicators) An example of how the above properties are useful. Let $X \sim Binomial(n, p)$ for n positive integer and $0 \leqslant p \leqslant 1$ (n is the number of independent identical binary trials and p is the probability of success). We can write

$$X = \sum_{i=1}^{n} I_i$$

where I_i is the indicator that i^{th} trial is a success (i.e. $I_i \overset{\text{i.i.d.}}{\sim} Bernoulli(p)$). We have

$$EI_i = 1 \cdot p + 0 \cdot (1-p) = p.$$

Therefore,

$$EX = \sum_{i=1}^{n} EI_i = \sum_{i=1}^{n} p = np.$$

Theorem For a non-negative random variable X (i.e. $f(x) = 0$ for $x < 0$).

$$EX = \begin{cases} \int_0^\infty (1 - F(x))dx, & X \text{ continuous} \\ \sum_{x=0}^\infty (1 - F(x)), & X \text{ discrete} \end{cases}$$

Proof:
We prove the continuous case first,

$$\begin{aligned}
\int_0^\infty [1 - F(x)]\, dx &= \int_0^\infty [1 - \Pr(X \leqslant x)]\, dx \\
&= \int_0^\infty \Pr(X > x) dx \\
&= \int_0^\infty \int_{y=x}^\infty f_X(y) dy dx \\
&= \int_0^\infty \int_{x=0}^y f_X(y) dx dy \\
&= \int_0^\infty y f_X(y) dy \\
&= EX.
\end{aligned}$$

Then, for discrete case, we have

$$\begin{aligned}\sum_{x=0}^{\infty}(1-F(x)) &= \sum_{x=0}^{\infty}\Pr(X>x)\\ &= \sum_{x=0}^{\infty}\sum_{y=x+1}^{\infty}\Pr(X=y)\\ &= \sum_{y=1}^{\infty}\sum_{x=0}^{y-1}\Pr(X=y)\\ &= \sum_{y=1}^{\infty}y\Pr(X=y)\\ &= EX\end{aligned}$$

Moments

Example (Minimizing distance) The expected value of a random variable has another property, one that we can think of as relating to the interpretation of EX as a good guess at a value of X.

Suppose we measure the distance between a random variable X and a constant b by $(X-b)^2$. The closer b is to X, the smaller this quantity is. We can now determine the value of b that minimizes $E\left[(X-b)^2\right]$ and, hence, will provide us with a good predictor of X. (Note that it does no good to look for a value of b that minimizes $(X-b)^2$, since the answer would depend on X, making it a useless predictor of X.)

We could proceed with the minimization of $E(X-b)^2$ by using calculus, but there is a simpler method:

$$\begin{aligned}E(X-b)^2 &= E(X-EX+EX-b)^2\\ &= E\left[(X-EX)+(EX-b)\right]^2\\ &= E(X-EX)^2+(EX-b)^2+2E\left[(X-EX)(EX-b)\right],\end{aligned}$$

where we have expanded the square. Note that $E\left[(X-EX)(EX-b)\right]=(EX-b)E(X-EX)=0$, since $EX-b$ is constant and comes out of the expectation, $E(X-EX)=EX-EX=0$. This means

$$E(X-b)^2 = E(X-EX)^2+(EX-b)^2.$$

Such that $E(X-b)^2$ is minimized at $b=EX$. And $E(X-EX)^2$ is actually the variance of X ($VarX=E(X-EX)^2$).

The various moments of a distribution are an important class of expectations.

Definition For each integer n, the nth *moment* of X (or $F_X(x)$), μ_n', is

$$\mu_n' = EX^n.$$

The nth *central moment* of X, μ_n, is

$$\mu_n = E(X - \mu)^n,$$

where $\mu = \mu'_1 = EX$.

Notes:

- $\mu'_0 = EX^0 = 1$
- μ'_1 is the *mean*, usually denoted by μ.
- $\mu_0 = E(X - \mu)^0 = 1$
- $\mu_1 = 0$
- $\mu_2 = E(X - EX)^2$ is the *variance*
- $\mu_3 = E(X - EX)^3$ is related to the *skewness*.
- $\mu_4 = E(X - EX)^4$ is related to the *kurtosis*.

Definition The *variance* of a random variable X is its second central moment, $\text{Var}(X) = E\left[(X - EX)^2\right]$. The positive square root of $\text{Var}(X)$ is the *standard deviation* of X.

The variance gives a measure of the degree of spread of a distribution around its mean. Figure 16.1 shows a plot of two samples, one sample draws 100 numbers from a normal distribution with mean 0 and variance 1, $N(0, 1)$. The other sample draws 100 numbers from a normal distribution with mean 0 and variance 100, $N(0, 100)$.

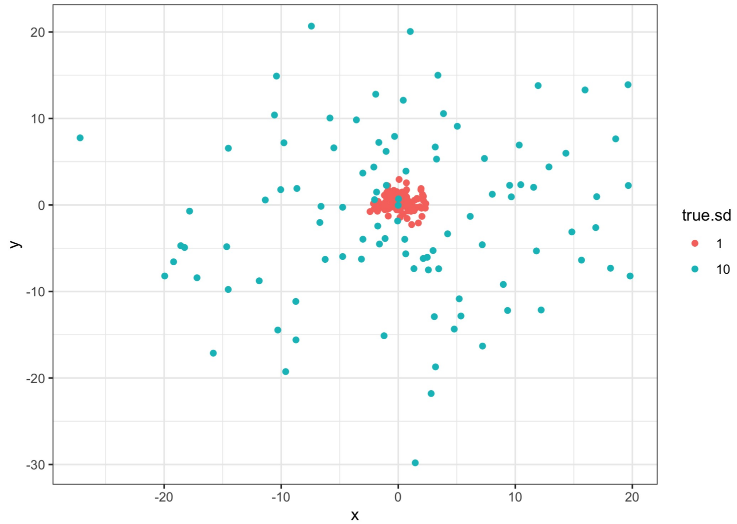

Figure 16.1: Figure 2.1.2. Two samples of 100 numbers drawn from $N(0, 1)$ and $N(0, 100)$.

Example (Exponential variance) Let X have the exponential(λ) distribution. We can calculate the variance of X now.

Solution:

$$\begin{aligned}\text{Var}\,(X) &= E(X-\lambda)^2\\ &= \int_0^\infty (x-\lambda)^2 \frac{1}{\lambda} e^{-x/\lambda} dx\end{aligned}$$

$$\begin{aligned}\text{Var}\,(X) &= \int_0^\infty (x^2 - 2x\lambda + \lambda^2) \frac{1}{\lambda} e^{-x/\lambda} dx\\ &= \int_0^\infty x^2 \frac{1}{\lambda} e^{-x/\lambda} dx - 2\int_0^\infty x\lambda \frac{1}{\lambda} e^{-x/\lambda} dx + \lambda^2\\ &= EX^2 - \lambda^2\\ &= \lambda^2\end{aligned}$$

Theorem If X is a random variable with finite variance, then for any constants a and b,

$$\text{Var}\,(aX+b) = a^2 \text{Var}\,(X).$$

Proof:
From the definition, we have

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

It is sometimes to use an alternative formula for the variance, given by

$$\text{Var}\,(X) = E(X^2) - (EX)^2,$$

which is easily established by

$$\begin{aligned}\text{Var}\,(X) &= E(X-EX)^2 = E\left[X^2 - 2XEX + (EX)^2\right]\\ &= EX^2 - 2(EX)^2 + (EX)^2\\ &= EX^2 - (EX)^2.\end{aligned}$$

Example (Binomial variance) Let $X \sim Binomial(n,p)$, that is ,

$$\Pr(X=x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

What is the variance of X?
Solutions:

Method #1:
We want to find EX^2 first. We use the

$$EX^2 = \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x}.$$

we use the same property $x^2\binom{n}{x} = xn\binom{n-1}{x-1}$. We then have

$$\begin{aligned}
EX^2 &= n \sum_{x=1}^{n} x \binom{n-1}{x-1} p^x (1-p)^{n-x} \\
&= n \sum_{y=0}^{n-1} (y+1) \binom{n-1}{y} p^{y+1} (1-p)^{n-1-y} \\
&= np \sum_{y=0}^{n-1} y \binom{n-1}{y} p^y (1-p)^{n-1-y} + np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} \\
&= np \cdot (n-1)p + np \\
&= n(n-1)p^2 + np.
\end{aligned}$$

And now

$$\begin{aligned}
\mathrm{Var}\,(X) &= EX^2 - (EX)^2 \\
&= n(n-1)p^2 + np - (np)^2 \\
&= np - np^2 \\
&= np(1-p).
\end{aligned}$$

Method #2:
Recall that we could write $X = \sum_{i=1}^{n} I_i$, where $I_i \overset{\text{i.i.d.}}{\sim} \mathit{Bernoulli}(p)$. Then

$$\begin{aligned}
\mathrm{Var}\,(X) &= \mathrm{Var}\left(\sum_{i=1}^{n} I_i\right) \\
&= \sum_{i=0}^{n} \mathrm{Var}\,(I_i) \qquad (I_i\text{'s are independent}) \\
&= n\mathrm{Var}\,(I_i) \qquad (I_i\text{'s are identically distributed}) \\
&= n\left[E(I_i^2) - (EI_i)^2\right] \\
&= n\left[p - p^2\right] \\
&= np(1-p).
\end{aligned}$$