MH2500 Probability and Introduction to Statistics

Handout 8 - Expected Values - IV

Synopsis

We introduce and study conditional expectations. We also viewed as random variables. Next, we introduce a very useful method when working with distributions, the moment generating function.

- Conditional Expectation and Prediction.
- Moment generating function.

Conditional Expectation and Prediction

We discuss the conditional mean associated to a conditional distribution.

Definition

Suppose X and Y are discrete random variables and that the conditional frequency function of Y given x is $p_{Y|X}(y|x)$. The **conditional expectation** of Y given $X = x$ is

$$E(Y|X = x) = \sum_y y p_{Y|X}(y|x).$$

For the continuous case, we have

$$E(Y|X = x) = \int y f_{Y|X}(y|x) dy.$$

More generally, the conditional expectation of a function $h(Y)$ is

$$E[h(Y)|X = x] = \int h(y) f_{Y|X}(y|x) dy$$

in the continuous case. A similar equation holds in the discrete case.

Example A

Consider a Poisson process on $[0, 1]$ with mean λ, and let N be the number of points in $[0, 1]$. For $p < 1$, let X be the number of points in $[0, p]$.

Find the conditional distribution and conditional mean of X given $N = n$.

. . . .

First, we find the joint distribution of $P(X = x, N = n)$, i.e., probability of x events in $[0, p]$ and $n - x$ events in $[p, 1]$.

From the assumption of a Poisson process, the count in the two intervals are independent Poisson random variables with parameters and

.

Hence

$$p_{X,N}(x, n) = \frac{(p\lambda)^x e^{-p\lambda}}{x!} \frac{[(1-p)\lambda]^{n-x} e^{-(1-p)\lambda}}{(n-x)!}.$$

Example A con't

The marginal distribution of N is Poisson, so the conditional frequency function of X is

$$\begin{aligned} p_{X|N}(x|n) &= \frac{1}{p_N(n)} p_{X,N}(x,n) \\ &= \frac{n!}{\lambda^n e^{-\lambda}} \frac{(p\lambda)^x e^{-p\lambda}}{x!} \frac{[(1-p)\lambda]^{n-x} e^{-(1-p)\lambda}}{(n-x)!} \\ &= \binom{n}{x} p^x (1-p)^{n-x}. \end{aligned}$$

This is binomial distribution with parameter n and p. From Section 4.1.2 (Handout 5), the conditional expectation is np.

. . . .

Note:

- $E(X|N=n) = np$. Hence $E(X|N)$ is a function of N and not X.
- In a similar way, for any two random variables X and Y, $E(Y|X)$ is a random variable that is a function of X.

Law of total expectation.

The mean of $E(Y|X)$ is given by $E[E(Y|X)]$.

Theorem A

$$E(Y) = E[E(Y|X)].$$

Theorem A might be called a **law of total expectation:**

The expectation of a random variable Y can be calculated by weighing the conditional expectation appropriately and summing or integrating.

Proof of Theorem A

We only prove the discrete case. The continuous case is similar.
Proof: First, by Section 4.1.1 Theorem A (Handout 5 Slide 14),

$$\begin{aligned} E[E(Y|X)] &= \sum_x E(Y|X=x)p_X(x) \\ &= \sum_x \left(\qquad\qquad\qquad \right) p_X(x). \end{aligned}$$

Changing the order of the summation gives

$$\begin{aligned} E[E(Y|X)] &= \sum_y \\ &= \sum_y y p_Y(y), \end{aligned}$$

by the law of total probability. Therefore, $E[E(Y|X)] = E(Y)$.

Example C

Suppose that in a system, a component and a backup unit both have mean lifetime equal to μ. If the component fails, the system automatically substitutes the backup unit, but there is a probability p that something will go wrong and it will fail to do so.
Let T be the total lifetime and let $X = 1$ if the substitution of the back up takes place successfully, and $X = 0$ if it does not. Find $E(T)$.

. . . .

The total lifetime is the lifetime of the first component if the backup fails, and is the sum of the original and backup unit if the backup is successfully made. Thus

$$E(T|X = 1) = 2\mu$$
$$E(T|X = 0) = \mu.$$

Therefore,

$$E(T) = E(T|X = 1)P(X = 1) + E(T|X = 0)P(X = 0) = \mu(2 - p).$$

Example D

Suppose the random variable N denotes the number of jobs in a single-server queue and X_i denotes the service time for the ith job. Let T denote the time to serve all jobs in the queue, i.e., $T = X_1 + X_2 + \cdots + X_N$. Prove that $E(T) = E(N)E(X)$. Does this agrees with intuition?

. . . .

By Theorem A,
$$E(T) = E[E(T|N)].$$

Since $E(T|N = n) = nE(X)$, we see that $E(T|N) = NE(X)$ and so
$$E(T) = E[NE(X)] = E(N)E(X).$$

This agrees with the intuitive guess that the average time to complete N jobs where N is random, is the average value of N times the average amount of time to complete a job.

Conditional variance

Recall that for a random variable Y the variance may be computed as

$$Var(Y) = E(Y^2) - [E(Y)]^2. \tag{$*$}$$

Now, in place of Y, we consider a conditional distribution $Y|X$.

For any value of x, we define

$$Var(Y|X = x) = E(Y^2|X = x) - [E(Y|X = x)]^2.$$

We may view $Var(Y|X)$ as a function of X and as a random variable. Since

$$Var(Y|X) = E(Y^2|X) - [E(Y|X)]^2,$$

we have

$$E[Var(Y|X)] = E[E(Y^2|X)] - E\{[E(Y|X)]^2\}.$$

and from $(*)$,

$$Var[E(Y|X)] = E\{[E(Y|X)]^2\} - \{E[E(Y|X)]\}^2.$$

Theorem B

Theorem B

$$Var(Y) = Var[E(Y|X)] + E[Var(Y|X)]$$

Proof:

By the law of total expectation, we see that

$$\begin{aligned} Var(Y) &= E(Y^2) - [E(Y)]^2 \\ &= E[E(Y^2|X)] - \{E[E(Y|X)]\}^2. \end{aligned}$$

Therefore, we have

$$\begin{aligned} Var(Y) &= E[E(Y^2|X)] - \{E[E(Y|X)]\}^2 \\ &= E[E(Y^2|X)] - E\{[E(Y|X)]^2\} + E\{[E(Y|X)]^2\} - \{E[E(Y|X)]\}^2 \\ &= \qquad\qquad\qquad\qquad\qquad . \end{aligned}$$

Example E - Random Sums

Suppose the number of insurance claim in a certain time period has expected value equal to 900 and standard deviation equal to 30, as would be the case if the number were a Poisson random variable with expected value 900. Suppose that the average claim value is \$1000 and the standard deviation is \$500.

Find $E(T)$ and variance of T.

. . . .

Like in the previous example,

$$E(T) = E(N)E(X) = 900 \times 1000 = 900,000.$$

By Theorem B,

$$Var(T) = Var[E(T|N)] + E[Var(T|N)].$$

Since $E(T|N) = NE(X)$,

$$Var[E(T|N)] = \qquad\qquad\qquad .$$

Example E con't

Also, since $Var(T|N=n) = Var(\sum_{i=1}^{n} X_i) = nVar(X)$,

$$Var(T|N) =$$

and

$$E[Var(T|N)] = \qquad\qquad .$$

Therefore

$$\begin{aligned} Var(T) &= Var[E(T|N)] + E[Var(T|N)] \\ &= [E(X)]^2 Var(N) + E(N)Var(X). \end{aligned}$$

Thus

$$\begin{aligned} Var(T) &= [1000]^2 \times 900 + 900 \times 500^2 \\ &= 1.125 \times 10^9. \end{aligned}$$

Moment-Generating Function

Here's a very useful tool that can dramatically simplify certain calculations.

The **moment-generating function (mgf)** of a random variable X is $M(t) = E(e^{tx})$ if the expectation is defined.

In the discrete case
$$M(t) = \sum_x e^{tx} p(x)$$

and in the continuous case,
$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$

- $M(t) = E(e^{tx})$ may or may not exists for any particular value of t.
- It depends on whether the tail of the series/ tail of the density decreases rapidly enough or not.

MGF

- E.g. for continuous case, the Cauchy density decreases at a rate of x^{-2} and so
$$M(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\pi(1+x^2)} dx \geq \int_{1}^{\infty} \frac{e^{tx}}{\pi x^2} dx$$
does not converge for any nonzero t.

Property A

The moment generating function exists for t in an open interval containing zero, it uniquely determines the probability distribution.

The proof depends on properties of the Laplace transform and so we cannot prove this important property here.

The derivative of $M(t)$ is
$$M'(t) = \frac{d}{dt} \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$

MGF

It can be shown that differentiation and integration can be interchanged, so that

$$M'(t) = \int_{-\infty}^{\infty} xe^{tx} f(x)dx \qquad \text{and} \qquad M'(0) = \int_{-\infty}^{\infty} xf(x)dx = E(x).$$

Differentiating r times, we find

$$M^{(r)}(0) = E(X^r).$$

It can be further argued that if the moment-generating function exists in an interval containing zero, then so do all the moments.

Property B

If the moment-generating function exists in an open interval containing zero, then $M^{(r)}(0) = E(X^r)$.

Example A - Poisson

Find the mgf of a Poisson distribution. Use it to find its variance.

$$M(t) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} \qquad = \qquad$$

$$= \qquad\qquad = \qquad .$$

The sum converges for all t. Differentiating, we have

$$M'(t) = \lambda e^t e^{\lambda(e^t - 1)}$$

$$M''(t) = \qquad\qquad .$$

Evaluating these derivatives at $t = 0$ gives

$$E(X) = \lambda, \qquad E(X^2) = \lambda^2 + \lambda,$$

and so

$$Var(X) = E(X^2) - [E(X)]^2 = \lambda.$$

Example B - Gamma

Find the mgf of a gamma distribution. Use it to find its variance.

. . . .

$$M(t) = \int_0^\infty e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$$

$$= \qquad\qquad\qquad\qquad .$$

This integral converges for $t < \lambda$ and can be evaluated by relating to the gamma density having parameters α and $\lambda - t$. Hence

$$M(t) = \qquad\qquad\qquad\qquad .$$

Example B - Gamma

Differentiating, we find that

$$M'(t) = \alpha \frac{\lambda^\alpha}{(\lambda - t)^{\alpha+1}}$$

$$M''(t) = \alpha(\alpha+1) \frac{\lambda^\alpha}{(\lambda - t)^{\alpha+2}}.$$

Evaluating these derivatives at $t = 0$ gives

$$E(X) = M'(0) =$$

$$E(X^2) = M''(0) = \qquad ,$$

and so

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{\alpha(\alpha+1)}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 \\ &= \frac{\alpha}{\lambda^2}. \end{aligned}$$

Example C - Standard Normal

Find the mgf of a standard normal distribution. Use it to find its variance.

. . . .

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx.$$

the integral converges for all t and can be evaluated using the technique of completing the square.

$$\frac{x^2}{2} - tx =$$

Therefore,

$$M(t) = \qquad\qquad\qquad .$$

Example C - Standard Normal

Making the change of variables $u = x - t$, and using the fact that the standard normal density integrates to 1, we find that

$$M(t) = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} du \quad = \qquad .$$

Therefore,

$$M'(t) = te^{t^2/2},$$
$$M''(t) = e^{t^2/2} + t^2 e^{t^2/2},$$

and

$$E(X) = M'(0) = 0, \qquad E(X^2) = M''(0) = 1.$$

Hence

$$Var(X) = 1.$$

Property C

If X has the mgf $M_x(t)$ and $Y = a + bX$, then Y has the mgf

$$M_Y(t) = e^{at} M_X(bt).$$

Proof:

$$\begin{aligned} M_Y(t) &= E(e^{tY}) \\ &= E(e^{at+btX}) \\ &= E(e^{at} e^{btX}) \\ &= e^{at} E(e^{btX}) \\ &= e^{at} M_X(bt). \end{aligned}$$

Example D - General Normal Distribution

Suppose $X \sim N(\mu, \sigma^2)$. Find the mgf of X and verify that $Var(X) = \sigma^2$.

. . . .

From Example C, we know that for $Z \sim N(0,1)$, we have $M_Z(t) = e^{t^2/2}$. Let $X = \mu + \sigma Z$. Then $X \sim N(\mu, \sigma^2)$. By Property C,

$$M_X(t) = \qquad\qquad = \qquad\qquad .$$

Therefore,

$$M'(t) = \qquad\qquad ,$$

$$M''(t) = \qquad\qquad ,$$

and

$$E(X) = M'(0) = \mu, \qquad E(X^2) = M''(0) = \sigma^2 + \mu^2.$$

Hence $Var(X) = \sigma^2$.

Property

Property D

If X and Y are independent random variables with mgf's M_X and M_Y and $Z = X + Y$, then $M_Z(t) = M_X(t)M_Y(t)$ on the common interval where both mgf's exist.

Proof:

$$\begin{aligned} M_Z(t) &= E(e^{tZ}) \\ &= E(e^{tX+tY}) \\ &= E(e^{tX}e^{tY}) \end{aligned}$$

by the assumption of independence,

$$\begin{aligned} &= E(e^{tX})E(e^{tY}) \\ &= M_X(t)M_Y(t). \end{aligned}$$

Example E – Independent Poisson

Suppose X is Poisson with parameter λ and Y is Poisson with parameter μ. Prove that $X+Y$ is Poisson with parameter $\mu+\lambda$.

From Example A, we know that X and Y have mgf

$$M_X(t) = e^{\lambda(e^t-1)} \qquad M_Y(t) = e^{\mu(e^t-1)}.$$

Hence

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{(\lambda+\mu)(e^t-1)}.$$

Therefore, $X+Y$ is Poisson with parameter $\lambda+\mu$.

Example F – Independent Gamma

(a) Suppose X follows a gamma distribution with parameters α_1 and λ while Y follows a gamma distribution with parameters α_2 and λ, and both X and Y are independent. Find the distribution of $X + Y$.

(b) Suppose $X_1, X_2, \ldots, X_n$ are independent and have exponential distributions with parameter λ. What is the distribution of $Z = X_1 + X_2 + \cdots + X_n$?

(a) From Example B, the mgf of $X + Y$ is

$$\left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1} \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_2} = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_1+\alpha_2}$$

where $t < \lambda$.
This mgf is recognized as the mgf of a gamma distribution with parameters λ and $\alpha_1 + \alpha_2$, which is the distribution of $X + Y$.

Example F – Independent Gamma

(b) Recall that an exponential distribution with parameter λ is a gamma distribution with parameter λ and 1.

. . . .

Recall that the time between one Poisson event to the next follows an exponential distribution. (Handout 2 slide 37).
From the above, the time between n consecutive Poisson events follow a gamma distribution.

Example G – Normal

Suppose $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$ and are independent of each other. Find the mgf of $X + Y$. What is its distribution?

....

By Example D,

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\mu t + \sigma^2 t^2/2} e^{\nu t + \tau^2 t^2/2} = e^{(\mu+\nu)t + (\sigma^2+\tau^2)t^2/2}.$$

Thus $X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2)$.

Remarks on mgf

1. In the past three examples, we have $X+Y$ having the same distribution as X and Y. This is not true in general. For example, if X is gamma with parameters λ_1 and α_1 and Y is gamma with parameters λ_2 and α_2. The the mgf of $X+Y$ is
$$\left(\frac{\lambda_1}{\lambda_1 - t}\right)^{\alpha_1}\left(\frac{\lambda_2}{\lambda_2 - t}\right)^{\alpha_2}$$
which is not necessarily gamma.

2. If X and Y have a joint distribution, their joint mgf is defined as
$$M_{XY}(s,t) = E(e^{sX+tY}).$$

3. It can be shown that two random variables X and Y are independent if and only if their joint mgf factors into the product of the mgf's of the marginal distributions.

4. A major limitation of the mgf is that it may not exist.