Properties of Expectation

We have already derived a few properties of expected values earlier in this course. Recall, that the expected value of a discrete random variable X with PMF $p(x)$ is defined as

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

whereas the expected value of a continuous random variable X with density $f(x)$ is defined as

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

Recall further, that we have shown that expectation is a linear operator. That is

$$E[aX + b] = aE[X] + b$$

Fact: If a random variable takes only values in a specific interval (i.e., $P(a \leq X \leq b) = 1$) then the expected value of X must be within the same interval.

Proof:

We will next take a look at how to compute expected values for functions of jointly distributed random variables.

Fact: Let X and Y be jointly distributed random variables with probability mass function $p(x, y)$ (if X and Y are discrete) or with probability density function $f(x, y)$ (if X and Y are continuous). Then

$$E[g(X, Y)] = \sum_x \sum_y g(x, y)p(x, y)$$

or

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)\, dx\, dy$$

Proof: Recall, that we have previously shown that for non-negative random variables

$$E[X] = \int_0^{\infty} P(X > t)dt$$

Similarly, one can show that if $g(X, Y) \geq 0$

$$E[g(X,Y)] = \int_0^\infty P\left(g(X,Y) > t\right) dt$$

Example 77. An accident occurs at a point that is uniformly distributed along a road of length L. At the time of the accident, an ambulance is at location Y also contributed uniformly along the same road. Find the expected distance of the ambulance from the accident assuming that X and Y are independent.

Fact: The expectation of a sum is always the same as the sum of the expectations.

$$E[X + Y] = E[X] + E[Y]$$

Note: For the above statement to hold, we *do not* have to assume independence.

Example 78. Mean of a hypergeometric random variable: If $X \sim \text{Hypergeometric}(N, m, n)$ show that $E[X] = \frac{nm}{N}$.

Example 79. Matching Problem

Consider once more the n people who all toss a personal item into a pile, turn off the lights and each select an item at random. Let X denote the number of people who get their own item back. Find $E[X]$.

Example 80. Coupon collector problem

Suppose that there are N different coupons and that each time one obtains a coupon it is equally likely to be any one of the N types. Find the expected number of coupons one has to collect in order to obtain a complete set.

Moments of the Number of Events that Occur

In some of the previous examples we were interested in finding the expected number of events $A_1, \ldots, A_n$ that occurred. The strategy for finding this expected value was to define indicator random variables

$$\mathbb{1}_{A_i} = \begin{cases} 1 & A_i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

Then

$$X = \sum_{i=1}^{n} \mathbb{1}_{A_i}$$

and

$$E[X] = E\left[\sum_{i=1}^{n} \mathbb{1}_{A_i}\right] = \sum_{i=1}^{n} E[\mathbb{1}_{A_i}] = \sum_{i=1}^{n} P(A_i)$$

Now, suppose instead that we are interested in the number of *pairs* of events that occurs. Since $\mathbb{1}_{A_i}\mathbb{1}_{A_j}$ is equal to one only if both indicators are equal to one, it follows that the number of pairs is equal to $\sum_{i<j} \mathbb{1}_{A_i}\mathbb{1}_{A_j}$. Since X is the number of events that occur it also follows that the number of pairs is $\binom{X}{2}$. Hence

$$\binom{X}{2} = \frac{X!}{(X-2)!2!} = \frac{X(X-1)}{2} = \sum_{i<j} \mathbb{1}_{A_i}\mathbb{1}_{A_j}$$

Taking expectations yields

$$E\left[\binom{X}{2}\right] = E\left[\frac{X(X-1)}{2}\right] = \sum_{i<j} E[\mathbb{1}_{A_i}\mathbb{1}_{A_j}] = \sum_{i<j} P(A_iA_j)$$

More generally, for some integer $k \leq n$

$$E\left[\binom{X}{k}\right] = \sum_{i_1<i_2<\cdots<i_k} E[\mathbb{1}_{A_{i_1}}\mathbb{1}_{A_{i_2}}\cdots\mathbb{1}_{A_{i_k}}] = \sum_{i_1<i_2<\cdots<i_k} P(A_{i_1}A_{i_2}\cdots A_{i_k})$$

Definition: Let X be a continuous (or discrete) random variable with density function $f(x)$ (or probability mass function $p(x)$). Then the k^{th} moment of X is defined as

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

or

$$E[X^k] = \sum_{x:p(x)>0} x^k p(x)$$

Example 81. Moments of a Binomial random variable

Let $X \sim \text{Binomial}(n, p)$. Derive the first three moments of X.

Covariance, Variance of Sums, and Correlations

Fact: If X and Y are independent, then for any functions h and g we have

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Proof:

Recall, that variance is a measure for how much observations deviate from the mean on average. It measures how different the observations are that are made on a single random variable. For jointly distributed random variables we are interested in a measure that describes how the random variables vary together.

Definition: The covariance between X and Y, denoted by Cov(X, Y) is defined by

$$Cov(X,Y) = E\left[(X - E[X])(Y - E[Y])\right] = E[XY] - E[X]E[Y]$$

Fact: If two random variables X and Y are independent, then their covariance is zero (Cov(X,Y) = 0).

Caution: The reverse is not true! Consider the joint PMF

		X		
		-1	0	1
Y	0	0	$1/3$	0
	1	$1/3$	0	$1/3$

Find $Cov(X, Y)$. Are X and Y independent?

Example 82. * Let X and Y denote the values of two stocks at the end of a five-year period. X is uniformly distributed on the interval (0,12). Given $X = x$, Y is uniformly distributed on the interval $(0, x)$. Calculate $Cov(X, Y)$.

We will next list some of the facts about covariances.

Fact: $Cov(X, Y) = Cov(Y, X)$, that is covariance is symmetric.

Proof:

Fact: $Cov(X, X) = Var(X)$.

Proof:

Fact: $Cov(aX + b, Y) = aCov(X, Y)$.

Proof:

Fact: $Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$.

Proof:

The above results allow us to also make a statement about the variance of a sum of (not necessarily independent) random variables.

Fact: For any random variables $X_1, \ldots, x_n$ it is

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$$

Proof:

Example 83. Find Var$(X + Y)$.

Fact: For pairwise independent random variables $X_i, \ldots, X_n$ it is

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

Example 84. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables each with mean μ and variance σ^2. Define

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

to be the sample average and sample variance, respectively. Find $Var(\bar{X})$ and $E[S^2]$.

Conditional Expectation and Variance

Recall, that if X and Y are discrete jointly distributed random variables then the conditional probability mass function of X given $Y = y$ is

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p(x,y)}{p_Y(y)}$$

This allows us to define the conditional expectation or variance of a random variable.

Definition: Let X and Y and be jointly distributed random variables with joint PMF $p(x, y)$ (or joint PDF $f(x, y)$). Then the conditional expectation of X given that $Y = y$ is defined as

$$E[X|Y = y] = \sum_x x p_{X|Y}(x|y) \quad \text{or} \quad E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$$

More generally, for any function $g(X)$ we have

$$E[g(X)|Y = y] = \sum_x g(x) p_{X|Y}(x|y) \quad \text{or} \quad E[g(X)|Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y) dx$$

The conditional variance of X given $Y = y$ is defined as

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

Example 85. * The stock prices of two companies at the end of any given year are modeled with random variables X and Y that follow a distribution with joint density function

$$f(x,y) = \begin{cases} 2x & 0 < x < 1,\ x < y < x + 1 \\ 0 & \text{otherwise} \end{cases}$$

Determine the conditional variance of Y given that $X = x$.

Moment Generating Functions

Recall, that the k^{th} moment of random variable X is defined as $E[X^k]$.

Definition: The moment generating function $M(t)$ of random variable X is defined as

$$M(t) = E[e^{tX}] = \begin{cases} \sum\limits_x e^{tx} p(x) & X \text{ is discrete} \\ \int\limits_{-\infty}^{\infty} e^{tx} f(x) dx & X \text{ is continuous} \end{cases}$$

The moments of X can be obtained by successively differentiating $M(t)$ and evaluating the result at $t = 0$. For example,

$$M'(t) = \frac{d}{dt} E[e^{tX}] = E\left[\frac{d}{dt} e^{tX}\right] = E[Xe^{tX}]$$

Hence,

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

Similarly,

$$M''(0) = E[X^2]$$

etc., so that for $k \geq 1$

$$M^{(k)}(0) = E[X^k]$$

Example 86. Find the moment generating function of the Poisson distribution and use it to derive mean and variance of this distribution.

Example 87. Find the moment generating function of the exponential distribution and use it to find the mean and variance of this distribution.

Fact: Suppose that X and Y are independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$, respectively. Then the moment generating function of $X+Y$ is given by

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

That is, the moment generating function of the sum of two independent random variables is the product of the moment generating functions.

Proof:

Below, find a table with moment generating functions of some common distributions.

Name	$M(t)$	Mean	Variance
Binomial(n,p)	$(pe^t+1-p)^n$	np	$np(1-p)$
Poisson(λ)	$exp(\lambda(e^t-1))$	λ	λ
Geometric(p)	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial(r,p)			
Uniform(a,b)	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential (λ)	$\frac{\lambda}{\lambda-t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma(r,λ)			
Normal(μ,σ^2)	$\exp\left(\mu t+\frac{\sigma^2t^2}{2}\right)$	μ	σ^2

Example 88. Find the moment generating function of a χ^2 random variable with n degrees of freedom.

Example*: An actuary determines that the claim size for a certain class of accidents is a random variable, X, with moment generating function

$$M_X(t) = \frac{1}{(1-2500t)^4}$$

Calculate the standard deviation of the claim size for this class of accidents.

Multivariate Normal Distribution

Recall, that the joint density of a bivariate normal distribution was

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ is the mean vector and Σ is the covariance matrix of the two random variables X_1 and X_2.

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_1, X_2) & Var(X_2) \end{pmatrix}$$

Now, instead of two random variables, consider a vector of n random variables

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)'$$

We say that the random vector $\mathbf{X}$ has a multivariate Normal distribution if it has the multivariate density

$$f(\mathbf{x}) = f(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}}|\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

Here

- $\boldsymbol{\mu}$ is the mean vector

$$\boldsymbol{\mu} = E(\mathbf{x}) = (\mu_1, \mu_2, \ldots, \mu_n)' = (E(X_1), E(X_2), \ldots, E(X_n))'$$

- Σ is the covariance matrix

$$\Sigma = E\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'\right] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}$$

where $\sigma_i^2 = Var(X_i)$ and $\sigma_{ij} = Cov(X_i, X_j)$. Note, that variances are always non-negative, but covariances can be either positive or negative. Covariance matrices are always symmetric (why?) and positive definite.

Alternative Definitions: The following definitions are equivalent to the PDF definition above.

(a) The random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ has a multivariate Normal distribution if every linear combination

$$Y = a_1 X_1 + \cdots + a_n X_n$$

is normally distributed.

(b) The random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ has a multivariate Normal distribution, if there is a random vector of independent standard normal random variables

$$\mathbf{Z} = (Z_1, Z_2, \ldots, Z_q)'$$

a fixed vector $\mu = (\mu_1, \ldots, \mu_n)'$ and a $n \times q$ matrix A such that

$$\mathbf{X} = A\mathbf{Z} + \mu$$

In this case $\Sigma = AA'$ is the covariance matrix of $\mathbf{X}$.

The Joint Distribution of the Sample Mean and Sample Variance

A while ago (when we discussed the t distribution) we were interested in finding the distribution of the t-test statistic. The result derived then was dependent on the fact, that for an IID normal sample the sample mean and sample variance are independent. We are now in a position to prove that fact.

Fact: Let $X_1, \ldots, X_n$ denote an independent sample from a Normal population with mean μ and variance σ^2. Further, define the sample mean

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

and sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Then

(i) $\bar{X} \sim \text{Normal}(\mu, \sigma^2/n)$

(ii) $\bar{X}$ and S^2 are independent.

(iii) $(n-1)S^2/\sigma^2 \sim \chi^2(\text{df = n-1})$

Proof: