All of Statistic
3 - Expectation

Levi Kassel

June 6, 2023

Summary

Expectation:

$$X \text{ is defined to be } E(X) = \int x dF(x) = \begin{cases} \sum_x x \cdot f(x) & \text{if } X \text{ is discrete} \\ \int x \cdot f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Note: To ensure that E(X) is well defined, we say that E(X) exists if

$$\int |x| dF_X(x) < \infty.$$

Otherwise, we say that the expectation does not exist.

Theorem (The Rule of the Lazy Statistician): Let $Y = r(X)$. Then

$$E(Y) = E(r(X)) = \int r(x) dF_X(x).$$

Moment: The kth moment of X is defined to be $E(X^k)$, assuming that $E(|X|^k) < \infty$.
Theorem: If the kth moment exists and if $j < k$, then the jth moment exists.

Theorem: If $X_1, \ldots, X_n$ are random variables and $a_1, \ldots, a_n$ are constants, then

$$E\left(\sum_i a_i X_i\right) = \sum_i a_i E(X_i).$$

Variance: Let X be a random variable with mean μ. The variance of X, denoted by σ^2, is defined by

$$\sigma^2 = E(X - \mu)^2 = \int (x - \mu)^2 \, dF(x)$$

Theorem: Assuming the variance is well defined, it has the following properties:
1. $V(X) = E(X^2) - \mu^2$.

2. If a and b are constants, then $V(aX+b) = a^2V(X)$.
3. If $X_1, \ldots, X_n$ are independent and $a_1, \ldots, a_n$ are constants, then

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V(X_i).$$

Sample Mean and Sample Variance: if $X_1, \ldots, X_n$ are random variables the sample mean is defined as:

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

And the sample variance is defined as:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Theorem: Let $X_1, \ldots, X_n$ be independent and identically distributed (iid) random variables, and let $\mu = E(X_i)$, $\sigma^2 = V(X_i)$. Then,

$$E(\overline{X}_n) = \mu, \quad V(\overline{X}_n) = \frac{\sigma^2}{n}, \quad \text{and} \quad E(S_n^2) = \sigma^2.$$

Covariance and Correlation: Let X and Y be random variables with means μ_X and μ_Y and standard deviations σ_X and σ_Y. The covariance between X and Y is defined as:

$$\text{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X)E(Y).$$

and the correlation is defined as:

$$\rho = \rho_{X,Y} = \rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

$$-1 \le \rho(X,Y) \le 1.$$

If $Y = aX + b$ for some constants a and b, then $\rho(X,Y) = 1$ if $a > 0$ and $\rho(X,Y) = -1$ if $a < 0$.
If X and Y are independent, then $\text{Cov}(X,Y) = \rho = 0$. However, the converse is not true in general.

Theorem: For random variables X and Y,

$$V(X+Y) = V(X) + V(Y) + 2\text{Cov}(X,Y)$$

and

$$V(X-Y) = V(X) + V(Y) - 2\text{Cov}(X,Y).$$

More generally, for random variables $X_1, \ldots, X_n$,

$$V\left(\sum_i a_i X_i\right) = \sum_i a_i^2 V(X_i) + 2\sum_{i<j} a_i a_j \text{Cov}(X_i, X_j).$$

Expectation and Variance of Important Random Variables

Distribution	Mean	Variance
Point mass at a	a	0
Bernoulli(p)	p	$p(1-p)$
Binomial(n, p)	np	$np(1-p)$
Geometric(p)	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson(λ)	λ	λ
Uniform(a, b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal(μ, σ^2)	μ	σ^2
Exponential(β)	β	β^2
Gamma(α, β)	$\alpha\beta$	$\alpha\beta^2$
Beta(α, β)	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
t_ν	0 (if $\nu > 1$)	$\frac{\nu}{\nu-2}$ (if $\nu > 2$)
χ^2_p	p	$2p$
Multinomial(n, p)	np	See below
Multivariate Normal(μ, Σ)	μ	Σ

Where if

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} = \begin{pmatrix} X_1 & X_2 & \cdots & X_k \end{pmatrix}$$

So:

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_k) \end{pmatrix}$$

$$\Sigma = \begin{bmatrix} V(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_k) \\ \text{Cov}(X_2, X_1) & V(X_2) & \cdots & \text{Cov}(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_k, X_1) & \text{Cov}(X_k, X_2) & \cdots & V(X_k) \end{bmatrix}$$

Lemma If a is a vector and X is a random vector with mean μ and variance Σ, then

$$E(a^T X) = a^T \mu \quad \text{and} \quad V(a^T X) = a^T \Sigma a.$$

If A is a matrix, then

$$E(AX) = A\mu \quad \text{and} \quad V(AX) = A\Sigma A^T.$$

Definition: The conditional expectation of X given $Y = y$ is defined as:

$$E(X|Y = y) = \begin{cases} \sum_x x f_{X|Y}(x|y) & \text{discrete case} \\ \int x f_{X|Y}(x|y)\, dx & \text{continuous case} \end{cases}$$

If $r(x,y)$ is a function of x and y, then the conditional expectation of $r(X,Y)$ given $Y=y$ is defined as:

$$E(r(X,Y)|Y=y) = \begin{cases} \sum_x r(x,y) f_{X|Y}(x|y) & \text{discrete case} \\ \int r(x,y) f_{X|Y}(x|y)\,dx & \text{continuous case} \end{cases}$$

Theorem (The Rule of Iterated Expectations): For random variables X and Y, assuming the expectations exist, we have that

$$E[E(Y|X)] = E(Y) \quad \text{and} \quad E[E(X|Y)] = E(X)$$

More generally, for any function $r(x,y)$, we have

$$E[E(r(X,Y)|X)] = E(r(X,Y))$$

Definition: The conditional variance is defined as

$$V(Y|X=x) = \int (y-\mu(x))^2 f(y|x)\,dy$$

where $\mu(x) = E(Y|X=x)$.

Theorem 3.27: For random variables X and Y,

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

Definition: The moment generating function (MGF) or Laplace transform of X is defined by

$$\psi_X(t) = E(e^{tX}) = \int e^{tx}\,dF(x)$$

where t varies over the real numbers.

$$\begin{aligned}
\psi'(0) &= \frac{d}{dt}\psi_X(t)\Big|_{t=0} \\
&= \frac{d}{dt}E(e^{tX})\Big|_{t=0} \\
&= E\left(\frac{d}{dt}e^{tX}\right)\Big|_{t=0} \\
&= E(Xe^{tX})\big|_{t=0} \\
&= E(X).
\end{aligned}$$

Therefore, $\psi'(0) = E(X)$.

$$\psi^{(k)}(0) = \frac{d^k}{dt^k}\psi_X(t)\Big|_{t=0} = E(X^k).$$

Properties of the MGF:

(1) If $Y = aX + b$, then $\psi_Y(t) = e^{bt}\psi_X(at)$.

(2) If $X_1, \ldots, X_n$ are independent and $Y = \prod_i X_i$, then $\psi_Y(t) = \prod_i \psi_i(t)$, where ψ_i is the MGF of X_i.

Theorem: Let X and Y be random variables. If $\psi_X(t) = \psi_Y(t)$ for all t in an open interval around 0, then X and Y are identically distributed.

Distribution	MGF, $\psi(t)$
Bernoulli(p)	$pe^t + (1-p)$
Binomial(n, p)	$(pe^t + (1-p))^n$
Poisson(λ)	$e^{\lambda(e^t-1)}$
Normal(μ, σ)	$e^{(\mu t + \frac{1}{2}\sigma^2 t^2)}$
Gamma(α, β)	$(1-\beta t)^{-\alpha}$ for $t < \frac{1}{\beta}$