Elements of Probability and Statistics

Lecture 10: Moments

IAI, TCG-CREST September 29, 2023

10.1 Summarizing distributions

We previously encountered two measures that are commonly used to summarize distributions. The *mean* or the expected value provides a measure of central tendency, i.e., the location where the center of the distribution lies. A measure of the spread of the distribution around its mean is provided by the *variance*. Some other measures of central tendency are the following.

(i) Median: h is the median of the distribution of a random variable X if $P(X \leq h) \geq 1/2$ and $P(X \geq h) \geq 1/2$, for all x.

(ii) Mode: For a *discrete* random variable, m is the mode if $P(X = m) \geq P(X = x)$ for all x. For a *continuous* random variable, m is the mode if it maximizes the PDF f, i.e., if $f(m) \geq P(x)$ for all x.

Moments: A family of summarizing measures including the mean and the variance fall under the study of *moments* of distributions. There are three kinds of moments, for a random variable X following a distribution with mean μ and variance σ^2.

(i) the n-th moment of X is $E(X^n)$.

(ii) the n-th central moment of X is $E((X - \mu)^n)$.

(iii) the n-th standardized moment of X is $E(\left(\frac{X-\mu}{\sigma}\right)^n)$.

One commonly used moment is the third standardized moment. used to measure the asymmetry of a distribution.

Skewness: $\text{Skew}(X) = E\left(\frac{X-\mu}{\sigma}\right)^3$.

By standardizing, the definition avoids depending on the location and scale of X. Another advantage is that the measure does not depend on the units (metres, kilograms, etc.) by which X is measured.

To see how skewness (or any odd standardized moment greater than or equal to 3) can be used as a measure of how distributions are asymmetric, we study the properties of symmetric distributions.

Symmetric Distribution: A random variable X has a symmetric distribution about μ if $X - \mu$ and $\mu - X$ have the same distribution.

We can note that this μ for a symmetric distribution is both the mean and the median of the distribution, which can be shown in the following way.

To show μ is mean: $E(X) - \mu = E(X - \mu) = E(\mu - X) = \mu - E(X) \implies E(X) = \mu$.

To show μ is median: $P(X - \mu \leq 0) = P(\mu - X \leq 0) \implies P(X \leq \mu) = P(X \geq \mu)$

$\implies P(X \leq \mu) = 1 - P(X > \mu) \geq 1 - P(X \geq \mu) = 1 - P(\leq \mu)$

$\implies P(X \le \mu) \ge 1/2$ and $P(X \ge \mu) \ge 1/2$

One property of interest for symmetric distributions is discussed next.

Thm. If X has symmetric distribution about μ, then in terms of the PDF of X, $f(x) = f(2\mu - x)$ for all x.

Proof. Let F be the CDF of X. Then,

$$F(x) = P(X - \mu \le x - \mu) = P(\mu - X \le x - \mu) = P(X \ge 2\mu - x) = 1 - F(2\mu - x)$$

Taking derivatives on both sides, $f(x) = f(2\mu - x)$.

Finally, the following property is relevant for the discussion on skewness.

Thm. If X has a symmetric distribution about μ, then for any odd n, the n-th central moment is $E(X - \mu)^n$ is 0 if it exists.

Proof. As $X - \mu$ and $\mu - X$ have the same distribution, they have the same odd n-th moment if it exists, i.e., $E(X - \mu)^n = E(\mu - X)^n$.

Now let $Y = (X - \mu)^n$. Then, $(\mu - X)^n = (-1)^n Y = -Y$.

Thus, $E(Y) = E(X - \mu)^n = E(\mu - X)^n = -E(Y)$, and therefore $E(Y) = 0$.

The converse is not true, as there exists asymmetric distributions that have long tails on one side and short wide tails on the other side, leading to skewness of 0.

On the use of skewness: The following considerations of (i) symmetric distributions have 0 skewness, and (ii) non-zero skewness implies asymmetric distributions, can be used to reach the following conclusion: of skewness is non-zero, then the distribution is asymmetric.

Positive skewness occurs when the distribution has longer right tails, and negative skewness occurs when the distribution has longer left tails.

As the 1st standardized moment is always 0, any odd moment greater than or equal to 3 can be used as a measure of asymmetry. The $3rd$ moment is popular for its relative ease in calculation.

Another moment that is often used is the fourth moment, as a measure of the nature of the tails of a distribution, in comparison with a normal distribution.

Kurtosis: $\text{Kurt}(X) = E(\left(\frac{X-\mu}{\sigma}\right)^4) - 3$.

The subtraction of 3 causes the kurtosis of any normal distribution to be 0. Compared to a standard normal distribution, heavier tails have positive kurtosis, whereas tails that have sharper decreases have negative kurtosis.

10.2 Moment Generating Functions

In mathematics, generating functions are continuous functions that are created corresponding to sequences of numbers, so that the elements of the sequence can be obtained from it. In probability, a Moment Generating Function (MGF) for a distribution generates all moments of the distribution, if they exist. An MGF is unique for a distribution, which makes it another way of specifying a distribution, in addition to PMFs/PDFs and CDFs. MGFs can also make it easy to prove certain distribution properties.

Moment Generating Function (MGF): The MGF of a random variable X is defined as a continuous function,

$$M(t) = E(e^{tX}),$$

if it is finite on an open interval $(-a, a)$ containing 0. Otherwise the MGF is said to not exist.

Any valid MGF will have $M(0) = 1$.

Q. Find the MGF for $X \sim \text{Bern}(p)$.

The function e^{tX} takes value e^t with probability p, and 1 with probability q. Thus $M(t) = E(e^{tX}) = pe^t + q$. Here $M(t)$ is finite and thus defined for all real t.

Q. Find the MGF for $X \sim \text{Geom}(p)$.

$$M(t) = E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} q^k p = p \sum_{k=0}^{\infty} (qe^t)^k = \frac{p}{1 - qe^t},$$

for $qe^t < 1$, i.e., for t in $(-\infty, \log(1/q))$, which is an interval containing 0.

Q. Find the MGF for $U \sim \text{Unif}(p)$.

$$M(t) = E(e^{tU}) = \frac{1}{b-a} \int_a^b e^{tu} du = \frac{e^{tb} - e^{ta}}{t(b-a)},$$

for $t \neq 0$, and $M(0) = 1$.

Thm. From the MGF of X, the n-th moment of X can be obtained as the n-th derivative of the MGF at 0, i.e., $E(X^n) = M^{(n)}(0)$.

Proof. From the Taylor expansion of $M(t)$ at 0, we get,

$$M(t) = \sum_{n=0}^{\infty} M^{(n)}(0) \frac{t^n}{n!},$$

From the series expansion,

$$M(t) = E(e^{tX}) = E\left(\sum_{n=0}^{\infty} X^n \frac{t^n}{n!} \right),$$

and by interchanging expectation and infinite sum (which can be done under certain conditions),

$$M(t) = \sum_{n=0}^{\infty} E(X^n) \frac{t^n}{n!}.$$

We can match the coefficients to obtain,

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

Thm. If X and Y are independent, then the MGF of $X + Y$ is the product of individual MGFs:

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

Proof. In the next lecture, we'll show two results: (i) if X and Y are independent, then any function involving both variables can be factorized as $f(X,Y) = g(X)h(Y)$, and (ii) If X and Y are independent, then $E(XY) = E(X)E(Y)$. By using the result of (i), the approach to prove (ii) can be followed to show that if X and Y are independent, then $E(f(XY)) = E(g(X))E(h(Y))$, where $f(X,Y) = g(X)h(Y)$.

Here this result is used to state that $E(e^{t(X+Y)}) = E(e^{tX})E(e^{tY})$., which proves $M_{X+Y}(t) = M_X(t)M_Y(t)$.

Q. Find the MGF for $X \sim \text{Binom}(n,p)$.

The Binomial random variable X can be written as a sum of n Bernoulli random variables, each of which have MGF $pe^t + q$. Using the previous theorem, the MGF of X is,

$$M(t) = (pe^t + q)^n.$$

Prop. If X has MGF $M(t)$, then after location-scale transformation the MGF of $a + bX$ is $E(e^{t(a+bX)}) = e^{at}E(e^{btX}) = e^{at}M(bt)$.

Q. Find the MGF of a Normal random variable.

The MGF of a standard Normal Z is,

$$M_Z(t) = E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz.$$

Introducing $e^{t^2/2}$ to complete the square,

$$M_Z(t) = e^{t^2/2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-(z-t)^2/2}dz = e^{t^2/2}.$$

Thus the MGF of $X = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$ is,

$$M_X(t) = e^{\mu t}M_Z(\sigma t) = e^{\mu t}e^{(\sigma t)^2/2} = e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

There can exist distributions for which the MGF does not exist, but the distribution has moments that can be calculated. This is demonstrated by the following two examples.

Q. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = e^X$ is said to follow a Log-Normal distribution. Find the MGF of $Y \sim \mathcal{LN}(\mu, sigma^2)$, and $E(Y)$, $Var(Y)$.

For $Y = e^Z$ with $Z \sim \mathcal{N}(0,1)$, the MGF does not exist, which can be shown from the following,

$$E(e^{tY}) = E(e^{te^Z}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{te^Z - z^2/2}dz,$$

which diverges as for any $t > 0$, $te^Z - z^2/2$ goes to infinity with increasing z. As $E(e^{tY})$ is not finite in an open interval around 0, the MGF of Y does not exist. By a similar argument, we can show that for $Y = e^X$ with $X \sim \mathcal{N}(\mu, \sigma^2)$, the MGF diverges and hence does not exist.

However, the n-th moment does exist, and can be found as,

$$E(Y^n) = E(e^{nX}) = M_X(n) = e^{n\mu + \frac{1}{2}n^2\sigma^2}.$$

Thus $E(Y) = e^{\mu + \frac{1}{2}\sigma^2}$, and $Var(Y)$ can be shown to be $m^2(e^{\sigma^2} - 1)$.

Defn. The Weibull distribution is often utilized to model the lifetime of any entity with a finite lifetime, instead of the Exponential distribution, which being memoryless is not suitable to capture this finite nature.

The Weibull distribution is a generalization of the Exponential distribution, where if $X \sim \text{Expo}(\lambda)$, then $T = X^{1/\gamma} \sim \text{Wei}(\lambda, \gamma)$, with $\lambda, gamma > 0$. The PDF of T is,

$$f(t) = \gamma \lambda e^{-\lambda t^{\gamma}} t^{\gamma - 1}, \text{ for } t > 0.$$

Q. $X \sim \text{Expo}(\lambda)$ has $E(X^n) = \frac{n!}{\lambda^n}$, find the n-th moment and the MGF of the Weibull random variable $T = X^{1/\gamma} \sim \text{Wei}(\lambda, \gamma)$, with $\lambda = 1$ and $gamma = 1/3$.

The n-th moment: $E(T^n) = E(X^{3n}) = (3n)!$.

The MGF: $E(e^{tT}) = E(e^{tX^3}) = \sum_0^{\infty} e^{tx^3 - x} dx$,

and as $tx^3 - x > x$ for sufficiently large x, the above integral diverges in comparison with the divergent integral $\sum_0^{\infty} e^x dx$. Thus the MGF does not exist.

Using MGFs, one can show that the sum of independent Normal random variables is Normal. Cramer's Theorem, whose proof is currently out-of-scope, states that if X_1 and X_2 are independent and $X_1 + X_2$ is Normal, then X_1 and X_2 must be Normal. An easier to prove result is that if X_1 and X_2 are i.i.d. random variables with an MGF $M(t)$, and $X_1 + X_2$ is Normal, then X_1 and X_2 are Normal. This is shown by assuming without loss of generality that if $X_1, X_2 \sim \mathcal{N}(0, 1)$, then its MGF is e^{t^2}, and is also,

$$e^{t^2} = E(e^{t(X_1 + X_2)}) = E(e^{t(X_1)}) E(e^{t(X_2)}) = (M(t))^2,$$

and hence $M(t) = e^{t^2/4}$, which is the MGF of $\mathcal{N}(0, 1/2)$. Thus $X_1, X_2 \sim \mathcal{N}(0, 1/2)$.

Avisek Gupta
Postdoctoral Fellow
IAI, TCG CREST.