Topic 5: Moments of a distribution

Rohini Somanathan

Course 003, 2018

The median, mode and quantile function

We have defined the mean and variance of a distribution. Other summary measures are:

Definition (Median): *For any random variable X, a median of the distribution of X is defined as a point m such that $P(X \leq m) \geq \frac{1}{2}$ and $P(X \geq m) \geq \frac{1}{2}$*

Definition (Mode): *For a discrete r.v. X, we say that c is the mode of X if it maximizes the PMF: $P(X = c) \geq P(X = x) \forall x$. For a continuous r.v. X, c is a mode if it maximizes the PDF: $f(c) \geq f(x) \forall x$.*

Definition (Quantiles): *When the distribution function of a random variable X is continuous and one-to-one over the whole set of possible values of X, we call the function F^{-1} the quantile function of X. The value of $F^{-1}(p)$ is called the p^{th} quantile of X or the $100 * p^{th}$ percentile of X for each $0 < p < 1$.*

What quantile is the median? A distribution can have multiple medians and modes, but the multiple medians have to occur side by side, whereas modes can occur all over a distribution.

Example: $X \sim \text{Unif}[a, b]$, **so** $F(x) = \frac{x-a}{b-a}$ **over this interval and the p^{th} quantile is given by:**

$$x = pb + (1-p)a$$

.

Compute this for $p = .5, .25, .9, \ldots$ and think about how the p^{th} quantile is affected by the shape of the density.

Finding quantiles: examples

1. **Find the medians for each of the following distributions:**
   (a) $P(X = 1) = .1 \;\; P(X = 2) = .2 \;\; P(X = 3) = .3 \;\; P(X = 4) = .4$
   (b) $P(X = 1) = .1 \;\; P(X = 2) = .4 \;\; P(X = 3) = .3 \;\; P(X = 4) = .2$
   (c)
$$f(x) = \begin{cases} \frac{1}{2} & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } 2.5 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$
2. **The p.d.f of a random variable is given by:**
$$f(x) = \begin{cases} \frac{1}{8}x & \text{for } 0 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$
   Find the value of t such that $P(X \leq t) = \frac{1}{4}$ **and** $P(X \geq t) = \frac{1}{2}$
   (Answers: $2, \sqrt{8}$**)**

The MAE and MSE

Result: *Let* $\mathbf{m}$ *and* $\boldsymbol{\mu}$ *be the median and the mean of the distribution of* $\mathbf{X}$ *respectively, and let* d *be any other number. Then the value* $\mathbf{d}$ *that minimizes the mean absolute error is* $\mathbf{d} = \mathbf{m}$*:*

$$E(|X - m|) \leq E(|X - d|)$$

and the value of $\mathbf{d}$ *that minimizes the mean squared error is* $\mathbf{d} = \boldsymbol{\mu}$*:*

$$E(X - \mu)^2 \leq E(X - d)^2$$

Moments of a random variable

Moments are special types of expectations that characterize the shape of a distribution.

Definition (Moments): *Let* X *be a random variable with mean* μ *and variance* σ^2*. For any positive integer* k*, then the*

k^{th} *moment of* X *is the expectation* $E(X^k)$

k^{th} *central moment is* $E[(X-\mu)^k]$ *and*

k^{th} *standardized moment is* $E[(\frac{X-\mu}{\sigma})^k]$

Comments:

The mean is the first moment and the variance is the second central moment.

If the k^{th} moment exists, all lower order moments exist and for bounded random variables, all moments exist.

If the distribution of X is symmetric with respect to its mean μ, and the central moment exists for a given odd integer k, then it must be zero (positive and negative terms cancel).

Skewness **is defined as the third standardized moment**

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

Sample moments

If we have a sample of i.i.d random variables, a natural way to estimate a population mean is to take the sample mean, the same is true for other moments of a distribution

Definition (Sample moments): *Let $X_1, \dots X_n$ be i.i.d. random variables. The k^{th} sample moment is the random variable*

$$M_k = \frac{1}{n}\sum_{j=1}^{n} X_j^k.$$

The sample mean $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ is the first sample moment.

The unbiasedness of the sample mean follows from the linearity of expectations.

Result (Mean and variance of sample mean): *Let $X_1, \dots X_n$ be i.i.d. random variables with mean μ and variance σ^2. Then the sample mean $\bar{X}_n$ is unbiased for estimating μ:*

$$E(\bar{X}_n) = \mu$$

Since the X_i are independent, the variance of the sample mean is:

$$Var(\bar{X}_n) = \frac{1}{n^2}Var(X_1 + \cdots + X_n) = \frac{\sigma^2}{n}$$

Moment generating functions

Definition (Moment generating function or MGF): *The moment generating function (MGF) of a random variable* $\mathbf{X}$ *is* $\mathbf{M}(\mathbf{t}) = \mathbf{E}(\mathbf{e}^{\mathbf{t}\mathbf{X}})$, *as a function of* $\mathbf{t}$, *if this is finite on some open interval (-a,a) containing 0. Otherwise we say the MGF of* $\mathbf{X}$ *does not exist.*

Result (MGF determines the distribution): *If two random variables have the same MGF on some open interval (-a,a) containing 0, they have the same distribution.*

MGFs help us to

calculate moments

identify distributions

Moments via derivatives of the MGF

Result (Moments via derivatives of the MGF): *Given the MGF of X, we can obtain the k^{th} moment of X by evaluating the k^{th} derivative of the MGF at 0.*

Proof. The function e^x can be expressed as the sum of the series $1 + x + \frac{x^2}{2!} + \ldots$ and so e^{tx} can be expressed as the sum $1 + tx + \frac{t^2x^2}{2!} + \ldots$ and the expectation $E(e^{tx}) = \sum\limits_{x=0}^{\infty}(1 + tx + \frac{t^2x^2}{2!} + \ldots)f(x)$. If we differentiate this w.r.t t and then set $t = 0$, we're left with only the second term in parenthesis, so we have $\sum\limits_{x=0}^{\infty} xf(x)$ which is defined as the expectation of X. Similarly, if we differentiate twice, were left with $\sum\limits_{x=0}^{\infty} x^2f(x)$, which is the second moment. For continuous distributions, we replace the sum $\sum\limits_{x=0}^{\infty}$ with an integral. $\int\limits_0^{\infty}(\ldots)dx$ □

Consider $f(x) = e^{-x}I_{(0,\infty)}$ (what this distribution is called?)

$$M(t) = \int\limits_0^{\infty} e^{x(t-1)}dx = \frac{e^{x(t-1)}}{t-1}\Big|_0^{\infty} = 0 - \frac{1}{t-1} = \frac{1}{1-t} \text{ for } t < 1$$

Taking the derivative of this function with respect to t, we get $\psi'(t) = \frac{1}{(1-t)^2}$, and differentiating again, we get $\psi''(t) = \frac{2}{(1-t)^3}$.

Evaluating the first derivative at $t = 0$, we get $\mu = \frac{1}{(1-0)^2} = 1$.

The variance $\sigma^2 = \mu_2' - \mu^2 = 2(1-0)^{-3} - 1 = 1$.

Properties of MGFs

1. **MGF of location-scale transformations:** *Let* X *be a random variable for which the MGF is* M_1 *and consider the random variable* $Y = aX + b$*, where* a *and* b *are given constants. Let the MGF of* Y *be denoted by* M_2*. Then for any value of* t *such that* $M_1(t)$ *exists,*

$$M_2(t) = e^{bt}M_1(at)$$

   Example: **If** $f(x) = e^{-x}I_{(0,\infty)}$ **as in the above example, the MGF of the random variable** $Y = (X - 1) = \frac{e^{-t}}{1-t}$ **for** $t < 1$ **(using the first result above, setting** $a = 1$ **and** $b = -1$**) and if** $Y = 3 - 2X$**, the MGF of** Y **is given by** $\frac{e^{3t}}{1+2t}$ **for** $t > -\frac{1}{2}$

2. **MGF of a sum of independent r.v.s:** *Suppose that* $X_1, \dots, X_n$ *are* n *independent random variables and that* M_i *is the MGF of* X_i*. Let* $Y = X_1 + \dots + X_n$ *and the MGF of* Y *be given by* M*. Then for any value of t such that* $M_i(t)$ *exists for all* $i = 1, 2, \dots, n$*,*

$$M(t) = \prod_{i=1}^{n} M_i(t)$$

We will now turn to several important applications of this second property.

Bernoulli and Binomial MGFs

Consider n Bernoulli r.v.s X_i with parameter p

The MGF for each of the X_i variables is given by

$$e^t P(X_i = 1) + (1) P(X_i = 0) = pe^t + q.$$

Using the additive property of MGFs for independent random variables, we get the MGF for $X = X_1 + \dots X_n$ as

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

For two Binomial random variables each with parameters (n_1, p) and (n_2, p), the MGF of their sum is given by the product of the MGFs, $(pe^t + q)^{n_1+n_2}$

Geometric and Negative Binomial MGFs

The density of a Geometric r.v. is $f(x;p) = pq^x$ over all natural numbers x

the MGF is given by

$$E(e^{tX}) = p\sum_{x=0}^{\infty}(qe^t)^x = \frac{p}{1-qe^t} \text{ for } t < \log(\frac{1}{q})$$

We can use this function to get the mean and variance, $\mu = \frac{q}{p}$ and $\sigma^2 = \frac{q}{p^2}$

The negative binomial is just a sum of r geometric variables, and the MGF is therefore $(\frac{p}{1-qe^t})^r$ and the corresponding mean and variance is $\mu = \frac{rq}{p}$ and $\sigma^2 = \frac{rq}{p^2}$

The Poisson MGF

Recall the Poisson PMF:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

for $x = 0, 1, 2, \dots$

$$E(e^{tX}) = \sum_{x=0}^{\infty} \frac{e^{tx}e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} = e^{\lambda(e^t-1)}$$

(Using the result that the series $1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots$ **converges to** e^z**, here** $z = \lambda e^t$**)**

We see from the form of the above MGF that the sum of k independently distributed Poisson variables has a Poisson distribution with mean $\lambda_1 + \dots \lambda_k$**.**

The Gamma MGF

Recall the density function for Gamma(α, β):

$$f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} I_{(0,\infty)}(x)$$

The MGF is therefore:

$$\begin{aligned}
M_X(t) &= \int_0^\infty e^{tx} f(x)\,dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta-t)x}\,dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(\beta-t)^\alpha} \int_0^\infty y^{\alpha-1} e^{-y}\,dy \quad (\text{ where } y = (\beta-t)x) \\
&= \left(\frac{\beta}{\beta-t}\right)^\alpha
\end{aligned}$$

Since a χ^2_v distribution, is Gamma$(\frac{v}{2}, \frac{1}{2})$, $M_X(t) = \frac{1}{(1-2t)^{\frac{v}{2}}}$.

Gamma transformations

Result (Gamma additivity): *Let* $X_1, \ldots X_n$ *be independently distributed random variables with respective gamma densities* $Gamma(\alpha_i, \beta)$. *Then*

$$Y = \sum_{i=1}^{n} X_i \sim Gamma(\sum_{i=1}^{n} \alpha_i, \beta)$$

Proof: **The MGF of Y is the product of the individual MGFs, i.e.**

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} \left(\frac{\beta}{\beta - t}\right)^{\alpha_i} = \left(\frac{\beta}{\beta - t}\right)^{\sum\limits_{i=1}^{n} \alpha_i} \quad \text{for } t < \beta$$

The Normal MGF

Let's begin deriving the MGF of $Z \sim N(0,1)$:

$$\begin{aligned} M_Z(t) &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-(z-t)^2}{2}} dz \end{aligned}$$

We completed the square and the term inside the integral is a $N(t,1)$ PDF, so integrates to 1. Therefore:

$$M_Z(t) = e^{\frac{t^2}{2}}$$

Any r.v. $X \sim N(\mu, \sigma^2)$ can be written as $X = \mu + \sigma Z$, so we can now use the location-scale result for MGFs to obtain

$$M_X(t) = e^{\mu t} M_Z(\sigma t)$$

So for a Normal r.v. $X \sim (\mu, \sigma^2)$:

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

Practice obtaining the moments of the distribution by taking derivatives of this function.

Normal transformations

Result 1: *Let* $X \sim N(\mu, \sigma^2)$ *and* $Y = aX + b$*, where* a *and* b *are given constants and* $a \neq 0$*, then* Y *has a normal distribution with mean* $a\mu + b$ *and variance* $a^2\sigma^2$

Proof: **The MGF of Y can be expressed as** $M_Y(t) = e^{bt}e^{\mu a t + \frac{1}{2}\sigma^2 a^2 t^2} = e^{(a\mu+b)t + \frac{1}{2}(a\sigma)^2 t^2}$**. This is simply the MGF for a normal distribution with the mean** $a\mu + b$ **and variance** $a^2\sigma^2$

Result 2: *If* $X_1, \dots, X_k$ *are independent and* X_i *has a normal distribution with mean* μ_i *and variance* σ_i^2*, then* $Y = X_1 + \cdots + X_k$ *has a normal distribution with mean* $\mu_1 + \cdots + \mu_k$ *and variance* $\sigma_1^2 + \cdots + \sigma_k^2$.

Proof: **Write the MGF of Y as the product of the MGFs of the** X_i**'s and gather linear and squared terms separately to get the desired result.**

We can combine these two results to derive the distribution of sample mean:

Result 3: *Suppose that the random variables* $X_1, \dots, X_n$ *form a random sample from a normal distribution with mean* μ *and variance* σ^2*, and let* $\bar{X}_n$ *denote the sample mean. Then* $\bar{X}_n$ *has a normal distribution with mean* μ *and variance* $\frac{\sigma^2}{n}$.

Note: **We already knew the mean and variance of the sample mean, we now have its distribution when the sample is Normal. The CLT extends this to other distributions for large samples.**

Transformations of Normals to χ^2 distributions

Result 4 : *If $X \sim N(0,1)$, then $Y = X^2$ has a χ^2 distribution with one degree of freedom.*

Proof:

$$\begin{aligned}
M_Y(t) &= \int_{-\infty}^{\infty} e^{x^2 t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2(1-2t)} dx \\
&= \frac{1}{\sqrt{(1-2t)}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\frac{1}{\sqrt{(1-2t)}}} e^{-\frac{1}{2}(x\sqrt{(1-2t)})^2} dx \\
&= \frac{1}{\sqrt{(1-2t)}} \quad \text{for } t < \frac{1}{2}
\end{aligned}$$

(the integrand is a normal density with $\mu = 0$ and $\sigma^2 = \frac{1}{(1-2t)}$).

The MGF obtained is that of a χ^2 random variable with $\nu = 1$ since the χ^2 MGF is given by $M_X(t) = (1-2t)^{-\frac{\nu}{2}}$ for $t < \frac{1}{2}$.

Normals and χ^2 distributions...

Result 5 : *Let* $X_1, \dots X_n$ *be independent random variables with each* $X_i \sim N(0,1)$*, then* $Y = \sum_{i=1}^{n} X_i^2$ *has a* χ^2 *distribution with* n *degrees of freedom.*

Proof:

$$
\begin{aligned}
M_Y(t) &= \prod_{i=1}^{n} M_{X_i^2}(t) \\
&= \prod_{i=1}^{n} (1-2t)^{-\frac{1}{2}} \\
&= (1-2t)^{-\frac{n}{2}} \quad \text{for } t < \frac{1}{2}
\end{aligned}
$$

which is the MGF of a χ^2 random variable with $\nu = n$. The parameter ν is called the degrees of freedom.