Moments and moment generating function

Dr. Linh Nghiem

STAT3023

Consider a sample space Ω with a probability measure P. Let X be a random variable defined on this sample space.

function of x

- Any random variable X has a cumulative distribution function (cdf) $F_X(x) = P(X \leq x)$.

X : random variable
x : realized value $P(X = x)$

- Discrete random variables:

X gets realized values on a countable set.

pmf : $P(X = x) = P_X(x)$

finite $\{1, 2, 3, \cdots 10\}$

infinite $\{1, 2, 3, \cdots\}$

- Continuous random variables:

X gets realized value on an incountable set.

Eg: $X \sim \text{Uniform}(0, 1)$

pdf : $f_X(x)$ density function :

$f_X(x) = \frac{dF_X(x)}{dx}$; $\int_{-\infty}^{\infty} f(x)\,dx = 1$; $f(x) \geq 0$.

$P(a \leq X \leq b) = \int_a^b f(x)\,dx$

Moments

$E(|X|)$ finite $\Rightarrow E(X)$ finite.

For any random variable X and a function $g : \mathbb{R} \to \mathbb{R}$, define the **expectation** of $g(X)$ to be

$$E[g(X)] = \begin{cases} \sum_x g(x)\, p_X(x) & X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x) f_X(x)\, dx & X \text{ is continuous.} \end{cases}$$

weighted avg.

(provided the sum or the integral is finite)

Examples:

- rth moment: $g(x) = x^r$; $\mu_r = E(X^r)$; $\mu = E(X)$
- rth central moment: $g(x) = (x-\mu)^r$; $\mu_r' = E[(X-\mu)]^r$

$r=2$; $E[(X-\mu)^2] = Var(X)$
$= E(X^2) - \mu^2$
$= \mu_2 - \mu^2$

Moment generating functions

- Moment generating function: encoding the sequence of moments $\{E(X^r)\}$, $r = 1, 2 \ldots, \infty$ into the coefficients of a power series.

  * Generating function: $\{a_r\}$, $r = 1, \ldots, \infty$

  $h(t) = \sum_{r=1}^{\infty} a_r \cdot c_r \cdot t^r$

- Choose $g(x) = \exp(tx)$ $= e^{tx}$, then the moment generating function (mgf) of a random variable X is defined to be

$$M_X(t) = E\{g(X)\} = E\{\exp(tX)\},$$

function of t.

provided this expectation exists for t in some open interval containing zero.

- $M_X(t) = \begin{cases} \sum_x \exp(tx) \cdot P_X(x), & X \text{ is discrete} \\ \int_{-\infty}^{\infty} \exp(tx) \cdot f_X(x)\, dx, & X \text{ is continuous} \end{cases}$

- $M_X(0) = E(e^0) = 1$.

Find $\begin{cases} t_1 < 0 \\ t_2 > 0 \end{cases}$ s.t. $E[e^{t_1 X}]$ and $E[e^{t_2 X}]$ are finite.

$t_1 \quad 0 \quad t_2 \quad t$

Getting moments from mgf

Calculus: $e^a = \sum_{r=0}^{\infty} \frac{a^r}{r!}$

- $M_X(t) = E[e^{tX}] = E\left[\sum_{r=0}^{\infty} \frac{(tX)^r}{r!}\right]$ $\{a = tX$

$= \sum_{r=0}^{\infty} \underbrace{E[X^r]}_{\downarrow} \cdot \frac{t^r}{r!}$

$E(X^r)$ is the coefficient for $\frac{t^r}{r!}$

($M_X(t)$ is finite $\Rightarrow$ interchange the sum and expectation) = not proved

- $\underbrace{\frac{d}{dt} M_X(t)}$ $= \frac{d}{dt}[E e^{tX}] = E\left[\frac{d}{dt} e^{tX}\right]$ (interchange the order if $M_X(t)$ finite)

$M_X'(t) = E[X e^{tX}]$; $M_X''(t) = E(X^2 e^{tX})$

$\underline{t=0}$: $M_X'(0) = E(X)$; $M_X''(0) = E(X^2)$

$\vdots$

$$\underline{M_X^{(r)}(0) = E(X^r)}$$

Examples

* $X \sim \text{Uniform}(0,1)$; $f_X(x) = 1$ for $\underline{0 \le x \le 1}$

$$M_X(t) = \int_0^1 e^{tx} \cdot f_X(x)\,dx = \int_0^1 e^{tx}\,dx = \left.\frac{e^{tx}}{t}\right|_{x=0}^{x=1}$$

$$= \begin{cases} \dfrac{e^t - 1}{t} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

Check moments: $M_X(t) = \dfrac{e^t - 1}{t} = \left(\sum_{r=0}^{\infty} \dfrac{t^r}{r!} - 1\right) / t$

$$= \left(\cancel{1} + \sum_{r=1}^{\infty} \frac{t^r}{r!} \cancel{-1}\right) / t = \sum_{r=1}^{\infty} \frac{t^{r-1}}{r!}$$

$$= \sum_{r=0}^{\infty} \frac{t^r}{(r+1)!} = \sum_{r=0} \frac{E[X^r] \cdot t^r}{r!}$$

$$\Rightarrow \frac{E[X^r]}{r!} = \frac{1}{(r+1)!} \Rightarrow E(X^r) = \frac{r!}{(r+1)!} = \frac{1}{r+1}$$

$E(X) = 1/2$;

$E(X^3) = 3/4$; ...

* $X \sim \text{Bern}(p)$: $P(X=1) = p$

$P(X=0) = 1-p$;

$$M_X(t) = E[e^{tX}] = e^{t \cdot 0}(1-p) + e^{t \cdot 1} p = 1 - p + pe^{t}$$

Examples

* $X \sim \text{Binomial}(n, p)$

$p_X(x) = P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots n$

$M_X(t) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \cdot e^{tx}$

$= \underbrace{\sum_{x=0}^{n} \binom{n}{x} [pe^t]^x (1-p)^{n-x}}$

$= [pe^t + (1-p)]^n;$

. Binomial thm

$\left[(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}\right]$

$a = pe^t; \quad b = (1-p)$

* $X \sim N(0,1); \quad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2};$

$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \cdot f_X(x)\, dx = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$

$tx - \frac{x^2}{2} = \frac{-1}{2}(x-t)^2 + \frac{1}{2}t^2$

$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx - \frac{x^2}{2}}\, dx;$

$= \underbrace{\left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2}\, dx\right]}_{\text{pdf of } N(t,1)} e^{t^2/2} = e^{t^2/2}$

$= 1$

$E(X^7)?$

$E(X^6)?$

* Check moments:

$M_X'(t) = t \cdot e^{t^2/2}; \quad M_X'(0) = 0 = E(X)$

$M_X''(t) = e^{t^2/2} + t^2 \cdot e^{t^2/2}; \quad M_X''(0) = 1 = E(X^2)$

$\text{Var}(X) = E(X^2) - [E(X)]^2$

$= 1 - 0 = 1.$

Uniqueness of mgf

" laplace transform .

- If the moment generating functions exists, and $M_X(t) = M_Y(t)$, then X and Y have the same distributions.
- Nevertheless, if two random variables have all the same moments, $E(X^r) = E(Y^r)$ for all $r = 1, 2, \dots$ then X and Y do not necessarily have the same distributions.

Example:

X ~ Std. log normal ; $f_X(x) = \frac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}\{\log(x)\}^2}, \; x > 0.$

tutorial sc

$M_X(t)$ does not exist

Y ~ $f_Y(y) = f_X(y)[1 + \sin(2\pi \log y)], \; y > 0,$

$E(X^r) = E(Y^r)$

$$E(Y^r) = \int_0^\infty y^r f_Y(y)\,dy = \underbrace{\int_0^\infty y^r f_X(y)\,dy}_{= E(X^r)} + \underbrace{\int_0^\infty y^r \frac{1}{y\sqrt{2\pi}} e^{-\frac{1}{2}\{\log y\}^2} \sin(2\pi \log y)\,dy}_{0}$$

$t = \log y - r; \quad dt = \frac{dy}{y}; \quad y = e^{t+r}$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{r(t+r)} \cdot e^{-\frac{1}{2}(t+r)^2} \cdot \sin(2\pi(t+r))\,dt$$

Properties of mgf

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{e^{-\frac{1}{2}(q^2 - r^2)} \cdot \sin(2\pi t)\, dt}_{\text{odd function}} = 0$$

Let X be a RV with mgf $M_X(t)$. Then the random variable $Z = aX + b$ has mgf

$$M_Z(t) = E[e^{tZ}] = E[e^{t(aX+b)}]$$
$$= E[e^{tb} \cdot e^{atX}]$$
$$= e^{tb} E[e^{atX}] = e^{tb} \cdot M_X(at)$$

$* \quad X \sim N(\mu, \sigma^2) \rightarrow X = \underset{a}{\sigma} Z + \underset{b}{\mu}; \quad Z \sim N(0,1)$

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

Properties of mgf

Recall for two **independent** random variables X and Y, we have

$$E\{g(X)h(Y)\} = E\{g(X)\}\,E\{h(Y)\}$$

for any two functions g and h. Let $M_X(t)$ and $M_Y(t)$ be mgfs of X and Y respectively, then the mgf of $Z = X + Y$ is given by

$$M_Z(t) = E[e^{Zt}] = E[e^{(X+Y)t}]$$
$$= \underbrace{E(e^{Xt} \cdot e^{Yt})}$$
$$= \underbrace{E[e^{Xt}]} \cdot E[e^{Yt}]$$

(X and Y are independent)

$$= M_X(t) \cdot M_Y(t)$$

Sum of independent random variables

More generally, if $X_1, \ldots, X_n$ be mutually independent random variables with mgfs $M_{X_i}(t)$ for $i = 1, \ldots, n$, then the mgf of $Z = \sum_{i=1}^{n} X_i$ is given by

$$M_Z(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

$$e^a = \sum_{x=0}^{\infty} \frac{a^x}{x!}$$

$$a = \lambda_i e^t$$

Eg: $X_1, \ldots X_n$; $X_i \sim \text{Poisson}(\lambda_i)$

$$Z = \sum_{i=1}^{n} X_i \; ?$$

$$M_Z(t) = \prod_{i=1}^{n} M_{X_i}(t) =$$

$$P(X_i = x) = e^{-\lambda_i} \lambda_i^x / x! \; ; \; x = 0, 1, \ldots \infty$$

$$M_{X_i}(t) = \sum_{x=0}^{\infty} e^{-\lambda_i} \cdot \frac{\lambda_i^x}{x!} e^{tx} = e^{-\lambda_i} \underbrace{\sum_{x=0}^{\infty} \frac{(\lambda_i e^t)^x}{x!}}$$

$$= e^{-\lambda_i} e^{\lambda_i e^t} = e^{\lambda_i (e^t - 1)}$$

Examples

$$M_Z(t) = \prod_{i=1}^{n} e^{\lambda_i(e^t - 1)} = \underbrace{e^{\left(\sum_{i=1}^{n}\lambda_i\right)(e^t-1)}}_{\text{MGF of Poisson}\left(\sum_{i=1}^{n}\lambda_i\right)}$$

$$\Rightarrow \quad Z \sim \text{Poisson}\left(\sum_{i=1}^{n}\lambda_i\right)$$

* $X_i \sim N(\mu_i, \sigma_i^2)$; $X_1, \dots X_n$: independent.

$Z = \sum_{i=1}^{n} X_i$?

$$M_Z(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

$$= \prod_{i=1}^{n} e^{\mu_i t + \frac{1}{2}\sigma_i^2 t^2} = \underbrace{e^{\left(\sum_{i=1}^{n}\mu_i\right)t + \frac{1}{2}\left(\sum_{i=1}^{n}\sigma_i^2\right)t^2}}_{\text{MGF of } N\left(\sum_{i=1}^{n}\mu_i,\ \sum_{i=1}^{n}\sigma_i^2\right)}$$

$$\Rightarrow Z \sim N\left(\sum_{i=1}^{n}\mu_i,\ \sum_{i=1}^{n}\sigma_i^2\right)$$

Probability bounds

Markov's inequality: For any non-negative random variable X and any $a > 0$, we have

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proof: X : continuous, $f_X(x)$

$$E(X) = \int_0^\infty x f_X(x)\,dx = \underbrace{\int_0^a x f_X(x)\,dx}_{\geq 0 \text{ (due to } x \geq 0)} + \int_a^\infty x f_X(x)\,dx$$

$$\geq \underbrace{\int_a^\infty x f_X(x)\,dx}_{} \geq \underbrace{\int_a^\infty a \cdot f_X(x)\,dx}_{= a \cdot \underbrace{\int_a^\infty f_X(x)\,dx}_{P(X \geq a)}}$$

$E(X) \geq a \cdot P(X \geq a)$

$\Rightarrow P(X \geq a) \leq \dfrac{E(X)}{a}$

Probability bounds

Chebyshev's inequality: For any random variable X and any $a > 0$, we have

$$P\left(|X - E(X)| \geq a\right) \leq \frac{\text{Var}(X)}{a^2}$$

Proof: $E(X) = \mu$;

$|X - E(X)| = |X - \mu| \geq 0$

$$P(|X-\mu| \geq a) = P(|X-\mu|^2 \geq a^2) \underset{\text{Markov}}{\leq} \frac{E|X-\mu|^2}{a^2} = \frac{E(X-\mu)^2}{a^2} = \frac{\text{Var}(X)}{a^2}$$

Probability bounds

Chernoff's bounds: for any RV X

$$P(X \geq a) = P(\underbrace{e^{tX}}_{\geq 0} \geq e^{ta}) \text{ for } t>0$$

$$\overset{\text{Markov}}{\leq} \frac{E[e^{tX}]}{e^{ta}} = \underbrace{e^{-ta} \cdot M_X(t)}_{h(t)}$$

This holds for any $t>0$, so we can find the lowest bound by minimizing $h(t)$

$$P(X \geq a) \leq \underbrace{\min_{t>0} e^{-ta} M_X(t)}_{\text{chernoff bound.}}$$

Examples

$X \sim Bin(n, p)$; $P(X \geq \alpha \cdot n)$ for $p < \alpha < 1$.

* Markov:

$$P(X \geq \alpha n) \leq \frac{E(X)}{\alpha n} = \frac{np}{\alpha n} = \frac{p}{\alpha} = \frac{2}{3}$$

. $p = 1/2$; $\alpha = 3/4$

* Chebyshev:

$$P(X \geq \alpha n) = P(\underbrace{X - np}_{} \geq \alpha n - np)$$

$$\leq P(|X - np| \geq \alpha n - np)$$

$$= \frac{Var(X)}{(\alpha n - np)^2} = \frac{np(1-p)}{n^2(\alpha - p)^2} = \frac{p(1-p)}{n(\alpha - p)^2}$$

$p = 1/2, \alpha = 3/4 \Rightarrow P(X \geq \alpha n) \leq 4/n$.

* Chernoff: $P(X \geq \alpha n) \leq \min_{t > 0} \underbrace{e^{-t\alpha n} \cdot \mathbb{E}[pe^t + (1-p)]^n}_{h(t)}$

Find t^* to minimize $h(t) = \{\underbrace{e^{-t\alpha}[pe^t + (1-p)]}_{g(t)}\}^n \Rightarrow t^*$ minimize $g(t)$

$g'(t) = 0 \to t^*$. $\underbrace{h(t^*) = \left[\frac{(1-p)}{(1-\alpha)}\right]^{n(1-\alpha)} \left[\frac{p}{\alpha}\right]^{\alpha n}}_{g(t^*)}$

Convergence of mgfs implies convergences of cdfs

$\alpha = 3/4$; $p = 1/2$; $h(t^*) = \left(\frac{16}{27}\right)^{n/4}$

Suppose $X_1, X_2, \ldots,$ is a sequence of random variables, each with mgf $M_{X_n}(t)$. Furthermore, suppose that

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$$

for all t in an open interval containing zero, and $M_X(t)$ is the mgf of a random variable X. Then for any x such that $F_X(x)$ is continuous, we have

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

We also say that the sequence $X_1, X_2, \ldots X_n$ converges to X in distribution.

Application: Poisson approximation to binomial distribution

Application: Central limit theorem

Convergence in probability

Weak law of large numbers