Lecture 17: Moment Generating Functions

$E(T|T>20) > E(T)$

how long someone's going to live, given that person lives to be at least 20 years

Human lifetimes are not memoryless. People get older and decay with age.

→ If human lifetimes were memoryless, ~~it would say~~ and the average is 80 years, it would say if we lived to be 20, then our new expectation is 100. Because memoryless says we are good as new. So, no matter how long we live, we get an extra 80 years on average, and that's not true empirically.

If memoryless, we would have $E(T|T>20) = 20 + E(T)$

→ It's realistic in problems where things don't decay with age.

Theorem: If X is a positive continuous random variable with memoryless property, then

$$X \sim \text{Expo}(\lambda), \text{ for some } \lambda.$$

memoryless property is ~~the~~ a property of the distribution, not of the random variable itself. (We would say the r.v. has the memoryless property if its distribution has the memoryless property.)

proof: Let F be the CDF of X, $G(x) = P(X > x)$
$= 1 - F(x)$

functional equation (solving for a function, not variable)

Memoryless property is
$G(s+t) = G(s)G(t)$. Solve for G

Not an usual equation where we're trying to solve for s or t or something like that. We are trying to solve for G, as we want to show that only exponential functions can satisfy this identity.

$s = t \Rightarrow G(2t) = G(t)^2,\ G(3t) = G(t)^3, \ldots,$
$G(kt) = G(t)^k$

$G(t/2) = G(t)^{1/2},\ G(t/3) = G(t)^{1/3}, \ldots, G(t/k) = G(t)^{1/k}$

$G(\frac{m}{n}t) = G(t)^{m/n}$. So, $\underbrace{G(xt) = G(t)^x}_{\text{by continuity}}$ for all real $x > 0$

G is continuous

$t = 1 \Rightarrow G(x) = G(1)^x = e^{x \underbrace{\ln G(1)}_{-\lambda\ (\lambda>0)}}$

prob. (between 0 and 1) ~~-ve~~ $[\ln(0-1) < 0]$

$= e^{-\lambda x}$

So, exponential is the only continuous memoryless distribution.

Moment Generating Function (MGF)

↳ MGF is another alternating way to describe a distribution, just like PDF, ~~CDF, etc.~~ and CDFs.

Definition : A random variable X has MGF
$M(t) = E(e^{tX})$, as a function of t,
if this is finite on some $(-a, a)$, $a > 0$.

t is just a placeholder → think of t as a book keeping device.
All the MGF is a fancy book keeping device for keeping track of the moments of a distribution.

Why is called Moment "generating"?

$$E(e^{tX}) = E\left(\sum_{n=0}^{\infty} \frac{X^n t^n}{n!}\right)$$
→ This is always valid because the Taylor series for e^x converges everywhere.

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

nth moment
1st moment is mean
1st & 2nd moments are used to compute ~~mean~~ variance

Why can we swap E and the summation?
→ If this were a finite sum, that would just be immediately true by linearity.
→ Since, it's an infinite sum, it requires more justification.
$(-a, a)$, $a > 0$ (S+a+210)

What we have done by bringing E inside the sum is just capture all the moments of X into the taylor series. That's why it's called the moment generating function because we see all the moments are just sitting there in the Taylor series.

Why is MGF important? (2) and (3) are important even if don't care about moments.

Let X has MGF $M(t)$.

(1) The nth moment, $E(X^n)$ is the coeff of $\frac{t^n}{n!}$ in the Taylor ~~series~~ of M and is ~~i.e.~~ $M^{(n)}(0)$ → nth derivative evaluated at 0

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

(2) MGF determines the distribution, i.e., if X, Y has same MGF, then they have the same CDF.

(3) In general, finding the distribution of a sum of independent random variables is complicated, that's called a convolution. But, if we have access to MGFs, things are a lot easier.

If X has MGF M_X, Y has MGF M_Y (M sub Y), X is independent of Y, then

MGF of $X+Y$ is $E(e^{t(X+Y)}) \underset{\text{(by independence)}}{=} E(e^{tX})E(e^{tY})$

$= M_X(t) M_Y(t)$

Example

$X \sim \text{Bern}(p)$, $M(t) = E(e^{tX}) = pe^{t \cdot 1} + q \underbrace{e^{t \cdot 0}}_{(1-p)}$

$= (pe^t + q)$

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

If we think of the binomial as the sum of i.i.d. Bern(p), then we use fact (3) there.

$$X \sim \text{Bin}(n,p) \Rightarrow M(t) = (pe^t + q)^n$$

$Z \sim N(0,1)$ → once we have the MGF of the standard normal, then we know the MGF of any normal we want. (location and scale)

MGF

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz - z^2/2}\, dz \quad \text{(by LOTUS)}$$

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

$$= e^{t^2/2}$$

Laplace Rule of Succession $X_1, X_2, \ldots$ i.i.d. Bern(p)

~~What's the probability that the sun be rise tomorrow?~~

Suppose that the sun has risen for the last n days in a row, and suppose we have observed and we have been alive for n days, every day the sun came up, n times in a row. What's the probability that the sun will rise tomorrow?

Given p, $\underbrace{X_1, X_2, \ldots \text{ i.i.d. Bern}(p)}_{\text{Conditional independence}}$

Known

Laplace is saying the probability that the sun rises is unknown. So, p is actually unknown. The question is how do we deal with that unknown p.

P unknown, Bayesian Approach: Treat p as a random variable and use Bayes' rule to find what's the distribution of p given all the evidence we have. We start with some prior beliefs about p, i.e. before we have any data or any evidence, we have some prior uncertainty. then we collect data and we use Baye's rule. → update based on evidence.

Laplace said:

Let $p \sim \text{Unif}(0,1)$ (prior)

Let $S_n = X_1 + \dots + X_n$

So, $S_n | p \sim \text{Bin}(n,p)$, $p \sim \text{Unif}(0,1)$

Conditional on p means p as known constant

Find posterior distribution. ~~(aft~~ (distribution after we collect the data)

$p | S_n$, and $P(X_{n+1} = 1 | S_n = n)$

$$f(p | S_n = k) = \frac{P(S_n = k | p) f(p)}{P(S_n = k)}$$

prior = 1

Uniform prior

Sun has risen for the last n days, what is the prob. that it would rise tomorrow.

pdf since p is continuous (Before we have data, we're treating as uniform. After we have data, it's going to have some density, a PDF, but it's conditional, so it's conditional pdf)

constant, does not depend on p (we're going to ignore the denominator because it doesn't depend on p)

$$P(S_n = k) = \int_0^1 P(S_n = k | p) f(p)\, dp$$

continuous version of Law of total probability

$$\propto p^k (1-p)^{n-k}$$

(also ignoring ^nC_k as it is a constant)

$$f(p | S_n = n) = (n+1) p^n$$

$$P(X_{n+1} = 1 | S_n = n) = \int_0^1 (n+1) p\, p^n\, dp$$

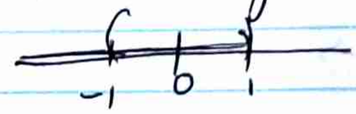
$$= \frac{(n+1)}{(n+2)}$$

So, a/c to Laplace, if the Sun rose 100 days in a row, then it would probably be $\left(\frac{101}{102}\right)$ for the next day

Lecture 18: MGFs Continued, Joint distributions

Expo MGF $X \sim \text{Expo}(1)$, find MGF, moments.

$$M(t) = E(e^{tX})$$
$$= \int_0^\infty e^{tx} e^{-x}\,dx \quad \text{(by LOTUS)}$$
$$= \int_0^\infty e^{-x(1-t)}\,dx$$
$$= \frac{1}{1-t}, \quad t < 1 \quad \text{(exponential decay, not exponential growth)}$$



$M'(0) = E(X)$, $M''(0) = E(X^2)$,
$M'''(0) = E(X^3), \ldots$

$$\frac{1}{1-t} = \sum_{n=0}^{\infty} t^n \quad [\,|t| < 1, \text{ Converges}\,]$$

n^{th} moment

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

$Y \sim \text{Expo}(\lambda) \rightarrow$ has mean $1/\lambda$
Let $X = \lambda Y \longrightarrow$ mean $= 1$
$\sim \text{Expo}(1)$

So, $Y^n = X^n/\lambda^n$

$$E(Y^n) = n!/\lambda^n$$

Example Let $Z \sim N(0,1)$, find all its moments.

$E(Z^n) = 0$ for n odd (by symmetry)

MGF $M(t) = e^{t^2/2}$

Unlike the geometric series, the Taylor series for e^x converges everywhere.

$$M(t) = e^{t^2/2} = \sum_{n=0}^{\infty} \frac{(t^2/2)^n}{n!} = \cancel{\tfrac{1}{2^k}} \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!} \frac{(2n)!}{(2n)!}$$

$$\Rightarrow E(Z^{2n}) = \text{Coeff. of } \frac{t^{2n}}{(2n)!}$$

$$= \frac{(2n)!}{2^n \cdot n!} \longrightarrow$$ number of ways to break $(2n)$ people into n ~~end to end~~ partnerships. (Practice 1,(2) 6)

$n = 1 \Rightarrow E(Z^2) = 1$
$n = 2 \Rightarrow E(Z^4) = 3$
$n = 3 \Rightarrow E(Z^6) = 1 \cdot 3 \cdot 5 = 15$

Example $X \sim \text{Pois}(\lambda)$

$$E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \lambda^k / k! \quad \text{(by LOTUS)}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} \longrightarrow$$ taylor series evaluated at $x = \lambda e^t$

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

$Y \sim \text{Pois}(\mu)$, X and Y are independent. Find distribution of $X + Y$.

Multiply MGFs: $e^{\lambda(e^t-1)} e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}$

$$\Longrightarrow X + Y \sim \text{Pois}(\lambda + \mu)$$

Sum of independent Poisson is still poisson.

Counterexample: if X, Y dependent:

$X = Y \Rightarrow X + Y = 2X$ is not Poisson, since even.

(I) A Poisson has to take on any possible non-negative integer value. Above is always an even number, so it couldn't possibly be a Poisson.

(II) $E(2X) = 2\lambda$, $Var(2X) = 4\lambda$ — for a Poisson, the mean always equals the variance.

(III) MGF's sum: $e^{(\lambda+\mu)(e^{2t}-1)}$

Joint Distributions (How do we work with the distribution of more than one random variable?)

X, Y Bernoulli (possibly dependent, possibly independent, possibly same p, possibly diff. p)

	$Y=0$	$Y=1$	
$X=0$	$2/6$	$1/6$	$3/6$
$X=1$	$2/6$	$1/6$	$3/6$
	$4/6$	$2/6$	

$\to \frac{1}{6} = \frac{3}{6} \times \frac{2}{6}$ They are independent.

2d discrete example

~~X, Y r.v.s, joint CDF, $F(x,y) = P(X \le x, Y \le y)$~~

~~joint PMF (discrete case), $P(X = x, Y = y)$~~

each of these joint probabilities is obtained by just multiplying two marginal probabilities.

$1/2$	0	$\frac{1}{2}$
$1/4$	$1/4$	$1/2$
$3/4$	$1/4$	

$\frac{1}{4} \times \frac{1}{2} \neq 0$ (so, not independent) thus dependent

X, Y random variables

Joint CDF

$F(x, y) = P(X \leq x, Y \leq y)$

Joint PMF (discrete case)

$P(X = x, Y = y)$

Marginal CDF

$P(X \leq x)$ is marginal distribution of X

Marginal Independence – joint ~~distribution~~ CDFs is the product of the marginal CDFs.

Joint PDF (Continuous case)

$f(x, y)$ such that

$$\underbrace{P((x, y) \in B)}_{\substack{(x,y) \text{ is in some set } B, \\ \text{where } B \text{ is some region} \\ \text{in the plane}}} = \iint_B f(x, y)\, dx\, dy$$

2D Case

Independence

~~Joint~~ X and Y are independent if and only if

$$\underbrace{F(x, y)}_{\text{joint CDF}} = \underbrace{F_X(x)}_{\text{marginal CDF of } x} \underbrace{F_Y(y)}_{\text{marginal CDF of } y}$$

Equivalent to

$$P(X=x, Y=y) = P(X=x)\,P(Y=y) \quad [\text{discrete case}]$$

marginal PMFs

$$\underbrace{f(x,y)}_{\text{joint PDF}} = \underbrace{f_X(x)\,f_Y(y)}_{\text{~~Joint~~ marginal PDFs}}$$

for all $x, y \in \mathbb{R}$

<u>Get marginal distributions from joint distributions</u>

$$P(X=x) = \sum_y P(X=x, Y=y) \quad (\text{Marginalizing over } Y)$$

(Discrete case)

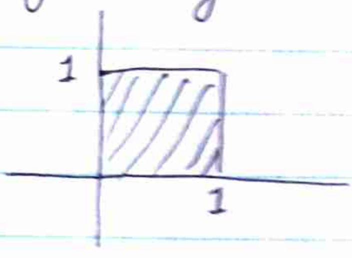
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

We can't go in the other direction. If we only know the marginal distributions, that doesn't tell us anything about how X and Y are related to each other.

<u>Example</u> Uniform on square $\{(x,y) : x, y \in [0,1]\}$

integral is area,

$$c = 1/\text{area} = 1$$

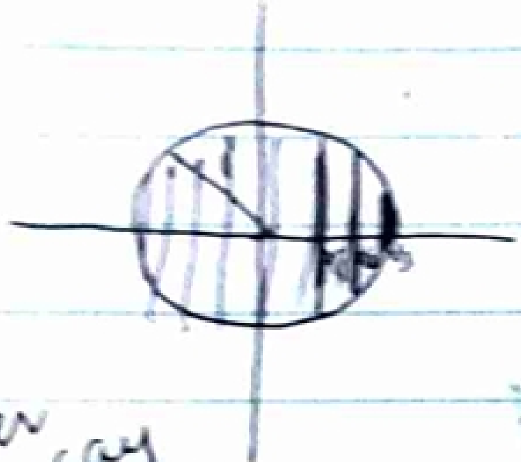


Marginally: X, Y are independent uniform $(0,1)$

joint PDF, const. on the square, 0 outside

$$\begin{cases} c, & 0 \le x \le 1, 0 \le y \le \\ 0, & \text{otherwise} \end{cases}$$

Example Unif in disc $x^2+y^2 \le 1$



joint PDF : $\begin{cases} 1/\pi, & x^2+y^2 \le 1 \\ 0, & \text{outside} \end{cases}$

X, Y dependent, because knowing X constraints the possible values of y.

Given $X=x$, $-\sqrt{1-x^2} \le y \le \sqrt{1-x^2}$,

Another way to say uniform is the prob. of some region must be proportional to it's area. In 1D, it's proportional to length.