MATH 180A: INTRO TO PROBABILITY
(FOR DATA SCIENCE)

www.math.ucsd.edu/~tkemp/180A

Today: REVIEW

Next: REVIEW

CAPEs! Please fill them out.

Lab 7 Due TODAY.

Homework 8 Due Friday, Dec 6.

Final Exam: Monday, Dec 9, 11:30a–2:30p in REC GYM

Extra/Extended Office Hours

Thursday 8:30 – 10:00am

Friday 10:00 – 11:30am

Main Topics Covered before Midterm 2

* Counting propability
* Conditional probability & Bayes' Rule
* Independence of Events
* Random variables & distributions
* CDF, PMF / PDF
* Independent trials & sampling
* $\mathbb{E}$ & Var
* Chebyshev's Inequality
* Ber, Bin, Geom, Poisson, $\mathcal{N}$, Exp, ...
* Normal Approximation & Confidence intervals
* Poisson Approximation
* ~~Poisson Process~~
* MGF

Moment Generating Function

Random Variable $X \rightsquigarrow$ MGF $M_X : \mathbb{R} \to (0, \infty]$

$$M_X(t) = \mathbb{E}(e^{tX})$$

$$= \sum_k e^{tk} \, \mathbb{P}(X=k) \quad \text{if } X \text{ is discrete}$$

$M_X(0) = 1$.

or

$$= \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx \quad \text{if } X \text{ is continuous \& has density } f_X.$$

PDF

$$\mathbb{E}(X^k) = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}$$

Some important examples: $X \sim \mathcal{N}(0,1)$, $Y \sim \mathrm{Exp}(\lambda)$, $N \sim \mathrm{Poisson}(\lambda)$

Theorem: If $M_X(t) < \infty$ for all $t \in$ some open interval including 0, then the function M_X uniquely determines the distribution of X.
I.e. If X, Y s.t. $M_X(t) = M_Y(t)$ for all $t \in$ a nbhd of 0, then $X \sim Y$.

$$M_X(t) = e^{t^2/2}$$

$$M_Y(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda \quad (\infty \text{ otherwise})$$

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

E.g. (5.17) $\quad f_X(x) = \begin{cases} x/2 & 0<x<2 \\ 0 & \text{otherwise} \end{cases}$

(a) Find $M_X(t)$. (b) Use it to compute $E(X^n)$ for all n.

(a) $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx = \int_0^2 e^{tx}\cdot \frac{x}{2}\,dx$

$$= \frac{1}{2}\int_0^2 \underbrace{x}_{u}\,\underbrace{e^{tx}\,dx}_{dv} = \frac{1}{2}\left(uv\Big|_0^2 - \int_0^2 v\,du\right) = \frac{1}{2}\left(\frac{x}{t}e^{tx}\Big|_{x=0}^{x=2} - \int_0^2 \frac{1}{t}e^{tx}\,dx\right)$$

$du = dx \quad v = \frac{1}{t}e^{tx}$

$$= \frac{2}{2t}e^{2t} - \frac{1}{2t}\int_0^2 e^{tx}\,dx$$

$$= \frac{1}{t}e^{2t} - \frac{1}{2t}\cdot\frac{1}{t}e^{tx}\Big|_0^2$$

$$= \frac{1}{t}e^{2t} - \frac{1}{2t^2}(e^{2t}-1)$$

$$= \frac{1}{2t^2}(2te^{2t} - e^{2t} + 1)$$

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

$$= \frac{1}{2t^2}\left(4t^2 + (2t-1)\left(\frac{(2t)^2}{2!} + \frac{(2t)^3}{3!} + \cdots\right)\right)$$

$$\cancel{= t^2 \times (\cdots)}$$

$$(2t-1)e^{2t} = (2t-1)\left(1 + 2t + \frac{(2t)^2}{2} + \frac{(2t)^3}{3!} + \cdots\right)$$

$$= \underbrace{2t - 1 + 2t(2t-1)} + (2t-1)\frac{(2t)^2}{2} + \cdots$$

$$= 4t^2 - \cancel{2t} + \cancel{2t} - 1$$

Eg. (5.18) Let $X \sim \text{Geom}(p)$. Compute $M_X(t)$. $\quad 0<p<1$

Use it to compute $\mathbb{E}(X)$ and $\text{Var}(X)$.

$$\sum_{k=1}^{\infty} x^k = \frac{x}{1-x}$$

$$\mathbb{P}(X=k) = (1-p)^{k-1}p. \qquad (k \geq 1)$$

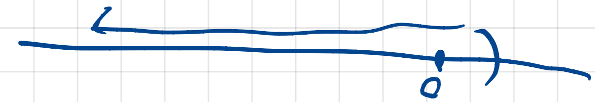
$$M_X(t) = \sum_{k=1}^{\infty} (1-p)^{k-1} p \cdot e^{tk} = \frac{p}{1-p}\sum_{k=1}^{\infty} (1-p)^k e^{tk}$$

$$(1-p)^{k-1} = \frac{(1-p)^k}{1-p}$$

$$= \frac{p}{1-p}\sum_{k=1}^{\infty} \left((1-p)e^t\right)^k$$

$(1-p)e^t < 1$.

i.e. $t < \ln\left(\frac{1}{1-p}\right)$



$$\frac{\cancel{p}}{\cancel{1-p}} \cdot \frac{\cancel{(1-p)}e^t}{1-(1-p)e^t} = \frac{pe^t}{1-(1-p)e^t}$$

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

$$\frac{d}{dt}\,\frac{p}{e^{-t}-(1-p)} = p\,\frac{d}{dt}\,\frac{1}{e^{-t}-(1-p)} = p \cdot \frac{-1}{(e^{-t}-(1-p))^2}\cdot(-e^{-t})$$

$$= \frac{pe^{-t}}{(e^{-t}-(1-p))^2}$$

$$\mathbb{E}(X) = M_X'(0)$$

$$= \frac{p}{(1-(1-p))^2} = \frac{p}{p^2} = \boxed{\frac{1}{p}}$$

MGF and Independent Sums

If $X_1, X_2, \ldots, X_n$ are independent, and $S_n = X_1 + \cdots + X_n$, then

$$M_{S_n}(t) = \mathbb{E}(e^{tS_n}) = \mathbb{E}\left(e^{t(X_1+X_2+\cdots+X_n)}\right) = \mathbb{E}(e^{tX_1} e^{tX_2} \cdots e^{tX_n})$$

independent

$$= \mathbb{E}(e^{tX_1})\mathbb{E}(e^{tX_2})\cdots\mathbb{E}(e^{tX_n}) = \boxed{M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t)}$$

E.g. Compute the MGF of a $\mathrm{Bin}(n,p)$ random variable. Use it to compute the mean and variance.

$$S_n \sim \mathrm{Bin}(n,p) \qquad \mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

$S_n \in \{0, 1, 2, \ldots, n\}$

$\mathbb{E}(S_n) = \sum_{j=1}^{n} \mathbb{E}(X_j) = np.$

$S_n = X_1 + X_2 + \cdots + X_n$, X_j indep. $\mathrm{Ber}(p)$

$\mathbb{P}(X_1 = 1) = p$

$\mathbb{P}(X_1 = 0) = 1-p.$

$\therefore \mathbb{E}(e^{tX_1})$

$= e^{t \cdot 1} p + e^{0 \cdot t}(1-p)$

$= \underline{1-p+pe^t}.$

$$\therefore M_{S_n}(t) = M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t) = M_{X_1}(t)^n$$

$$= (1-p+pe^t)^n.$$

$\mathbb{E}(S_n) = M_{S_n}'(0)$

$$\frac{d}{dt}(1-p+pe^t)^n = n(1-p+pe^t)^{n-1} \cdot pe^t. \Big|_{t=0} = n(1-p+p)^{n-1} \cdot p \cdot 1 = \boxed{np}$$

Eg. (8.13) Suppose $M_Z(t) = \left(\frac{1}{2}e^{-t} + \frac{2}{5} + \frac{1}{10}e^{t/2}\right)^{36} < \infty$ for all $t \in \mathbb{R}$

Show that Z can be written as a sum of i.i.d. random variables X_j, and describe the distribution of X_j.

$Z = X_1 + X_2 + \cdots + X_{36}$ where X_j i.i.d.

with $M_{X_j}(t) = \frac{1}{2}e^{-t} + \frac{2}{5}e^{0t} + \frac{1}{10}e^{t/2}$

$$= \sum_k e^{tk}\, \mathbb{P}(X_j = k)$$

$k \in \{-1, 0, 1/2\}$

$\mathbb{P}(X_j = -1) = \frac{1}{2}$

$\mathbb{P}(X_j = 0) = \frac{2}{5}$

$\mathbb{P}(X_j = \frac{1}{2}) = \frac{1}{10}$

$+$

1

E.g. Let $U \sim \text{Unif}[-1,1]$, $X \sim N(1,1)$, independent.

Find the MGF of $U+X$.

$$M_U(t) = \int_{-1}^{1} e^{tx} \frac{1}{1-(-1)}\,dx = \frac{1}{2}\int_{-1}^{1} e^{tx}\,dx = \frac{1}{2}\left(\frac{1}{t}e^{tx}\right)\Big|_{x=-1}^{1}$$

$$= \frac{1}{2t}\left(e^{t}-e^{-t}\right)$$

$X = Z+1$

$Z \sim N(0,1)$

$$M_X(t) = \mathbb{E}(e^{tX}) = \mathbb{E}(e^{t(Z+1)}) = \mathbb{E}(e^{tZ}e^{t})$$

$$= e^{t}\,\mathbb{E}(e^{tZ}) = e^{t}e^{t^2/2}.$$

$$\therefore\ M_{U+X}(t) = M_U(t)M_X(t) = \frac{1}{2t}(e^{t}-e^{-t})\cdot e^{t}\cdot e^{t^2/2}$$

$$= \frac{(e^{2t}-1)}{2t}\, e^{t^2/2}$$