Lecture - 9

MS Theory I

Review

Moment Generating Functions (Not just for generating moments!)

- Generate moments
- Characterize a dist'n (used to show convergence)

Definition 2.3.6 Let X be a random variable with cdf F_X. The *moment generating function (mgf) of* X (or F_X), denoted by $M_X(t)$, is

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

provided that the expectation exists for t in some neighborhood of 0. That is, there is an $h > 0$ such that, for all t in $-h < t < h$, $\mathrm{E}e^{tX}$ exists. If the expectation does not exist in a neighborhood of 0, we say that the moment generating function does not exist.

More explicitly, we can write the mgf of X as

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx \qquad \text{if } X \text{ is continuous,}$$

or

$$M_X(t) = \sum_x e^{tx} P(X = x) \qquad \text{if } X \text{ is discrete.}$$

It is very easy to see how the mgf generates moments. We summarize the result in the following theorem.

Theorem 2.3.7 *If X has mgf $M_X(t)$, then*

$$\mathrm{E}\, X^n = M_X^{(n)}(0),$$

← How to get moments from mgf. (derivatives easier than integrals).

where we define

$$M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\bigg|_{t=0}.$$

That is, the nth moment is equal to the nth derivative of $M_X(t)$ evaluated at $t = 0$.

Theorem 2.3.11 *Let $F_X(x)$ and $F_Y(y)$ be two cdfs all of whose moments exist.*

a. *If X and Y have bounded support, then $F_X(u) = F_Y(u)$ for all u if and only if $\mathrm{E}\, X^r = \mathrm{E}\, Y^r$ for all integers $r = 0, 1, 2, \ldots$.*

b. *If the moment generating functions exist and $M_X(t) = M_Y(t)$ for all t in some neighborhood of 0, then $F_X(u) = F_Y(u)$ for all u.*

Review

$M_X(t)$ exists $\forall t$
$-h<t<h \quad h>0$

many dist'ns

$$F_X \text{ (or } f_X) \longrightarrow M_X(t) \longrightarrow \mu_r' = E[X^r]$$

$r=0,1,2,\ldots$

bounded support

fewer dist'ns

Theorem 2.3.12 (Convergence of mgfs) *Suppose $\{X_i, i=1,2,\ldots\}$ is a sequence of random variables, each with mgf $M_{X_i}(t)$. Furthermore, suppose that*

$$\lim_{i\to\infty} M_{X_i}(t) = M_X(t), \qquad \text{for all } t \text{ in a neighborhood of } 0,$$

and $M_X(t)$ is an mgf. Then there is a unique cdf F_X whose moments are determined by $M_X(t)$ and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{i\to\infty} F_{X_i}(x) = F_X(x).$$

That is, convergence, for $|t|<h$, of mgfs to an mgf implies convergence of cdfs.

Summary of Thm 2.3.12

$$\lim_{i\to\infty} M_{X_i}(t) = M_X(t) \quad \forall |t|<h \quad \overset{\text{implies}}{\Longrightarrow} \quad \lim_{i\to\infty} F_{X_i}(x) = F_X(x)$$

$$\boxed{\text{mgfs converge} \Longrightarrow \text{cdfs converge}}$$

Review

The approximation: $P(X=x) \approx P(Y=x)$

for $X \sim bin(n,p)$ $\quad \left.\begin{matrix} Y \sim Pois(\lambda) \\ \lambda = np \end{matrix}\right]$ $\quad$ n 'large' np small $\quad \left\{\begin{matrix} \text{One rule:} \\ n \geq 20 \\ p \leq .05 \\ np < 10 \end{matrix}\right.$

Poisson (λ) pmf: $P(Y=y) = \frac{e^{-\lambda}\lambda}{y!} * I_{[0,1,2,\dots)}(y)$

Recall $\underset{\text{Binomial}}{M_X(t)} = [pe^t + (1-p)]^n$ and $\underset{\text{Poisson}}{M_Y(t)} = e^{\lambda(e^t-1)}$

If $\lambda = np$ then $p = \lambda/n$

$$M_X(t) = [pe^t + (1-p)]^n = \left[1 + \tfrac{1}{n}(e^t-1)(np)\right]^n$$

if $p = \frac{\lambda}{n}$ $\quad = \left[1 + \frac{1}{n}(e^t-1)\lambda\right]^n$ $\quad$ [Now use Lemma 2.3.14

define:
$a = a_n = (e^t-1)\lambda$

Then $\lim_{n\to\infty} M_X(t) = e^{\boxed{\lambda(e^t-1)}} = M_Y(t)$

$\therefore \lim_{n\to\infty} F_X(u) = F_Y(u) \quad u = [0,1,2,\dots)$

Note: we must recognize $e^{\lambda(e^t-1)}$ is Mgf of Poisson for this to be useful Thm.

Finally:

Theorem 2.3.15 *For any constants a and b, the mgf of the random variable $aX+b$ is given by*

$$M_{aX+b}(t) = e^{bt}M_X(at).$$

→ know dist'n and mgf of X

→ know mgf → dist'n of $Y = aX + b$
(linear transform).

MS Theory I

Lecture 9

4

Review

§2.4 Differentiating under integral sign

- We will frequently need to interchange intergration & differentiation.

- Need to know when legitimate to interchange $\int$ & $\frac{d}{d\theta}$

- Assume interested in calculating:

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)\,dx \qquad -\infty < a(\theta),\ b(\theta) < \infty$$

- Use application of Fundamental Thm of Calculus and chain rule.

Theorem 2.4.1 (Leibnitz's Rule) *If $f(x,\theta)$, $a(\theta)$, and $b(\theta)$ are differentiable with respect to θ, then*

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)\,dx = f(b(\theta),\theta)\frac{d}{d\theta}b(\theta) - f(a(\theta),\theta)\frac{d}{d\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta} f(x,\theta)\,dx.$$

Notice that if $a(\theta)$ and $b(\theta)$ are constant, we have a special case of Leibnitz's Rule:

partial derivative

$$\frac{d}{d\theta}\underbrace{\int_a^b f(x,\theta)\,dx}_{\text{ft'n of }\theta} = \int_a^b \underbrace{\frac{\partial}{\partial\theta} f(x,\theta)}_{\text{ft'n }x,\theta}\,dx.$$

Example Beta $(\alpha, 1)$

Beta(α,β)	
pdf	$f(x\mid\alpha,\beta) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, \quad 0\le x\le 1, \quad \alpha>0, \quad \beta>0$
mean and variance	$\mathrm{E}X = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{Var}\,X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
mgf	$M_X(t) = 1 + \sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^k}{k!}$
notes	The constant in the beta pdf can be defined in terms of gamma functions, $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. Equation (3.2.18) gives a general expression for the moments.

$$f(x\mid\alpha,\beta=1) = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(1)}x^{\alpha-1}(1-x)^{\beta-1} * I_{[0,1]}(x)$$

$$= \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)0!}x^{\alpha-1} * I_{[0,1)}(x) = \alpha x^{\alpha-1} I_{[0,1]}(x)$$

Review

$$f(x|\alpha, \beta=1) = \alpha x^{\alpha-1} I_{[0,1]}(x)$$

want $\frac{d}{d\alpha} \int_0^1 g(x,\alpha)\, \alpha x^{\alpha-1}\, dx$

$g(x,\alpha)$ often something like $\log(f(x|\alpha))$... (more later).

here $a(\theta)=1$ and $b(\theta)=0$ (both constant)! ☺

so $\frac{d}{d\alpha} \int_0^1 g(x,\alpha)\, \alpha x^{\alpha-1} dx = \int_0^1 \frac{\partial}{\partial\alpha}\left[g(x,\alpha)\, \alpha x^{\alpha-1}\right] dx$

Example $U(0,\theta)$

Caution:

Sample space is ft'n of parameters that define dist'n!

Uniform(a, b)

pdf	$f(x\|a,b) = \frac{1}{b-a}, \quad a \le x \le b$
mean and variance	$\mathrm{E}X = \frac{b+a}{2}, \quad \mathrm{Var}\, X = \frac{(b-a)^2}{12}$
mgf	$M_X(t) = \frac{e^{bt}-e^{at}}{(b-a)t}$
notes	If $a = 0$ and $b = 1$, this is a special case of the beta ($\alpha = \beta = 1$).

$$f(x) = \frac{1}{\theta} I_{[0,\theta]}(x) \qquad a(\theta)=0, \quad b(\theta)=\theta$$

$$\frac{d}{d\theta}\int_0^\theta g(x,\theta)\frac{1}{\theta}\,dx \neq \int_0^\theta \left[\frac{\partial}{\partial\theta} g(x,\theta)\frac{1}{\theta}\right] dx$$ ☹

$$\frac{d}{d\theta}\int_0^\theta g(x,\theta)\frac{1}{\theta}\,dx = \frac{g(\theta,\theta)}{\theta}\frac{d}{d\theta}\theta - \frac{g(0,\theta)}{\theta}\frac{d}{d\theta}(0) + \int_0^\theta \frac{\partial}{\partial\theta}\frac{g(x,\theta)}{\theta}\,dx$$

Review

More theorems in C&B to cover when $0<x<\infty$ or $-\infty<x<\infty$
$a(\theta)$ "=" $-\infty$ or $b(\theta)$ "=" ∞

Can we assume:

$$\frac{d}{d\theta}\int_{-\infty}^{\infty} f(x,\theta)\,dx = \int_{-\infty}^{\infty}\frac{\partial}{\partial\theta} f(x,\theta)\,dx \ ?$$

The property illustrated ~~for the exponential distribution~~ above holds for a large class of densities, which will be dealt with in Section 3.4.

'exponential families'
"obviously" the sample space / support won't be a ft'n of θ.

Similarly for 'exponential families' we can assume that we can interchange differentiation and summation.
$\frac{d}{d\theta}$ and $\sum_{i=1}^{\infty}$ or $\sum_{i=1}^{n}$.

More in chapter-3

Chapter - 3

Course Overview

C&B Chapters:

1: Probability Theory
2: Transformations / Expectations
3: Common Families of Distributions
4: Multiple Random Variables
5: Properties of a Random Sample
6: Principles of Data Reduction
7: Point Estimation
8: Hypothesis Testing
9: Interval Estimation
10: Asymptotic Evaluations

Foundation — Gaining Tools (chapters 1–5) — 1st Semester
Applying Tools (chapters 6–10) — 2nd Semester

Chapter - 3

3.1 Intro
3.2 Discrete Dist'ns
3.3 Continuous Dist'ns
3.4 Exponential families
3.5 Location and Scale families
3.6 Inequalities and Identities

- We've seen families of dist'n

  Binomial (n, p)
  Poisson (λ)
  Normal (μ, σ^2)
  Uniform (a, b)
  Beta (α, β)
  Gamma (α, β)
  Geometric (p)
  ⋮

  Note 'family' characterized by parameters.

  Know parameters, know how random variables behave
  ↳ See chapter 7 !

- For 'common' dist'ns:
  - $f_X(x \mid \theta)$ pdf "|" given parameter θ
  - $F_X(x \mid \theta)$
  - Sample space (support) of X. → $I(x)$
  - $E[X]$ = mean
  - $Var[X]$ = variance } ft'n of parameters
  - mgfs ← if exists / helpful

§3.2 Discrete Dist'ns.

- Sample Space of X countable.
- Usually

$$I_{[0,1,2,\dots)}(x) = \begin{cases} 1 & \text{if } x \in [0,1,2,\dots) \\ 0 & \text{else} \end{cases}$$

or

$$I_{[1,2,3,\dots)}(x) = \begin{cases} 1 & \text{if } x \in [1,2,3,\dots) \\ 0 & \text{else} \end{cases}$$

or

$$I_{[0,1,2,\dots n]}(x) = \begin{cases} 1 & \text{if } x \in [0,1,2,\dots n] \\ 0 & \text{else} \end{cases}$$

Usually... integers

Discrete Uniform Dist'n: (rolling standard die $N=6$)

Discrete uniform

get cdf by summing

pmf	$P(X=x\mid N)=\frac{1}{N};\quad x=1,2,\dots,N;\quad N=1,2,\dots$
mean and variance	$\mathrm{E}X=\frac{N+1}{2},\quad \mathrm{Var}\,X=\frac{(N+1)(N-1)}{12}$
mgf	$M_X(t)=\frac{1}{N}\sum_{i=1}^{N}e^{it}$

$$I_{[1,\dots N]}(x) = \begin{cases} 1 & x \in [1,\dots N] \\ 0 & \text{else} \end{cases}$$

$$E[X] = \sum_{i=1}^{N} x\, P(X=x\mid N)$$

$$= \sum_{i=1}^{N} x * \frac{1}{N}$$

$$= \frac{1}{N}\sum_{i=1}^{N} x = \frac{1}{\cancel{N}}\left(\frac{\cancel{N}(N+1)}{2}\right)$$

$$= \frac{N+1}{2}$$

Prove $\sum_{i=1}^{N} x = \frac{N(N+1)}{2}$ using proof by induction

Show $N=1$: $N=1 \quad \sum_{i=1}^{1} i = \frac{1+1}{2} = 1 \quad = \frac{1(1+1)}{2} = 1$

Assume $\sum_{i=1}^{k} i = \frac{k(k+1)}{2}$

Show $\sum_{i=1}^{k+1} i = \frac{(k+1)(k+2)}{2}$

$$\sum_{i=1}^{k} i + k+1 = \frac{k(k+1)}{2} + \frac{2(k+1)}{2} = \frac{(k+1)(k+2)}{2} \quad //$$

Discrete Uniform cont.

$$E[X] = \frac{N+1}{2}$$

$$Var[X] = E[X^2] - (E[X])^2$$

$$E[X^2] = \sum_{x=1}^{N} x^2 \cdot \frac{1}{N} = \frac{1}{N}\sum_{x=1}^{N} x^2 = \frac{(N+1)(2N+1)}{6}$$

Another proof by induction (homework).

$$Var(x) = E[x^2] - (E[x])^2$$

$$= \frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2$$

$\vdots$

$$= \frac{(N+1)(N-1)}{12}$$

→ Roll 6 sided die (N=6)

- pmf

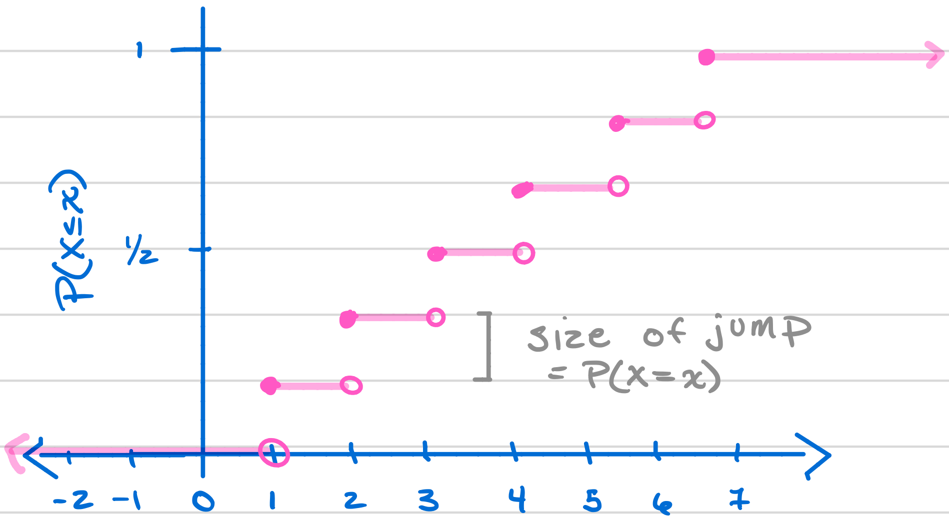
$$f_X(x \mid N=6) = P(X=x) = \frac{1}{6} * I_{[1,2,\dots 6]}^{(x)}$$

$$E[X] = \frac{6+1}{2} = 3.5$$

$$Var[X] = \frac{(7)(5)}{12} = \frac{35}{12} \cong 2.92$$

step function

- cdf $\underbrace{P(X \le x) = \sum_{i=1}^{x} P(X=i)}_{x \in [1,2,\dots N]}$



$\vdots$

size of jump $= P(X=x)$

$P(X<1) = 0$

$P(X \le 1) = P(X=1) = 1/6$

$P(X \le 2) = P(X=1) + P(X=2) = 2/6 = 1/3$

$\vdots$

$P(X \le 6) = P(X=1) + \dots + P(X=6) = 6/6 = 1$

→ Roll 3-sided die: $f_X(x) = \frac{1}{3} * I_{[0,1,2,3]}^{(x)}$

$E[X] = 2$ $\quad Var[X] = \frac{(4)(2)}{12} = \frac{8}{12} = \frac{2}{3}$

Hypergeometric

pmf	$P(X=x\mid N,M,K)=\frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}};\quad x=0,1,2,\ldots,K;$ $M-(N-K)\le x\le M;\quad N,M,K\ge 0$
mean and variance	$\mathrm{E}X=\frac{KM}{N},\quad \mathrm{Var}\,X=\frac{KM}{N}\frac{(N-M)(N-K)}{N(N-1)}$
notes	If $K \ll M$ and N, the range $x=0,1,2,\ldots,K$ will be appropriate.

- derivation of E[X] in C&B ← somewhat unsatisfying...
- Fun pmf ; Not used often in Biostats

Bernoulli(p)

Binomial with $n=1$

pmf	$P(X=x\mid p)=p^x(1-p)^{1-x};\quad x=0,1;\quad 0\le p\le 1$
mean and variance	$\mathrm{E}X=p,\quad \mathrm{Var}\,X=p(1-p)$
mgf	$M_X(t)=(1-p)+pe^t$

Binomial(n,p)

usually parameter n assumed known

pmf	$P(X=x\mid n,p)=\binom{n}{x}p^x(1-p)^{n-x};\quad x=0,1,2,\ldots,n;\quad 0\le p\le 1$
mean and variance	$\mathrm{E}X=np,\quad \mathrm{Var}\,X=np(1-p)$
mgf	$M_X(t)=[pe^t+(1-p)]^n$
notes	Related to Binomial Theorem (Theorem 3.2.2). The *multinomial* distribution (Definition 4.6.2) is a multivariate version of the binomial distribution.

Theorem 3.2.2 (Binomial Theorem) *For any real numbers x and y and integer $n\ge 0$,*

(3.2.4)
$$(x+y)^n=\sum_{i=0}^{n}\binom{n}{i}x^iy^{n-i}.$$

If $x=y=1$
$(1+1)^n=2^n=\sum_{i=0}^{n}\binom{n}{i}$

Poisson Dist'n

Poisson(λ)

pmf	$P(X = x\|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!};\quad x = 0, 1, \ldots;\quad 0 \le \lambda < \infty$
mean and variance	$\mathrm{E}X = \lambda, \quad \mathrm{Var}\, X = \lambda$
mgf	$M_X(t) = e^{\lambda(e^t - 1)}$

- Model wait times (time between hormone pulses).
- Number successes in given time (small time interval)
- Count data (often $E[X] = Var[X]$ not valid medical research).

$$f_X(x) = P(X=x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} * I_{[0,1,2,\ldots)}(x)$$

Recall:

$$\boxed{\text{Tool: } e^y = \sum_{i=0}^{\infty} y^i/i!}$$

Show pmf sums to 1.

$$\sum_{x=0}^{\infty} P(X=x|\lambda) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$$

$$E[X] = \sum_{x=\cancel{0}\,1}^{\infty} \frac{x e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty}\frac{\cancel{x} e^{-\lambda}\lambda\cdot\lambda^{x-1}}{\cancel{x}(x-1)} = \lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}$$

let $y = x-1$ $\quad x=1 \Rightarrow y=0$, $\quad x=\infty \Rightarrow y=\infty$

$$E[X] = \lambda e^{-\lambda}\sum_{y=0}^{\infty}\frac{\lambda^y}{y!} = \lambda \cancel{e^{-\lambda}}\cancel{e^{\lambda}} = \lambda$$

Similarly: $Var[X] = \lambda$

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

homework (after exam)

→ Hormone Pulse Example (waiting time)

- A hormone pulses on average 5 times in 3 hours
- Find probability there are no pulses in next hour.
- Find probability there are at least 2 pulses.

X = # pulses in hour   Possible X: $0,1,2,\ldots$

$X \sim \text{Poisson}(\lambda)$ λ ?

$\lambda = E[X] = 5 \text{ pulses}/3 \text{ hours} = \frac{5}{3}$

$$P(0 \text{ pulses in next hour}) = P(X=0) = \frac{e^{-5/3}\left(\frac{5}{3}\right)^0}{0!} = .189$$

$$P(\geq 2 \text{ calls in next hr}) = 1 - P(X=0) - P(X=1)$$

$$= 1 - .189 - \frac{e^{-5/3}\left(\frac{5}{3}\right)^1}{1!} = .496$$

//

→ Poisson Approx. to Binomial

Recall mgf of Binomial $\xrightarrow{\text{converges}}$ Mgf of Poisson] n large, np small

(Ex 2.3.13) $\Rightarrow$ cdfs converge

If $X \sim \text{Bin}(1500, \frac{1}{500})$ $P(X \leq 2) = \sum_{x=0}^{2} \binom{1500}{x}\left(\frac{1}{500}\right)^x\left(\frac{499}{500}\right)^{1500-x} = .4230$

easy in R; tough by hand...

$E[X] = np = \frac{1500}{500} = 3$

→ Poisson approx.

$$P(X \leq 2) \cong \sum_{x=0}^{2} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-3}\left[\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}\right] = .4232$$

Negative Bionomial

trials to r^{th} success, r fixed & known.

X = trial r^{th} success

r successes ; $(x-r)$ failures [note last must be success!]

→ in $(x-1)$ trials have $(r-1)$ successes

Success on r^{th} trial

$$p\left[\binom{x-1}{r-1} p^{(r-1)} (1-p)^{(x-1)-(r-1)}\right]$$

$I_{[r,r+1,r+2,\ldots)}(x) = \begin{cases} 1 & \text{if } x \in [r, r+1, \ldots) \\ 0 & \text{else} \end{cases}$

$$= \binom{x-1}{r-1} p^r (1-p)^{x-r} * I_{[r,r+1,r+2,\ldots)}(x)$$

Alternative form (homework)

Y = # failures to r^{th} success

r successes + y fail $\quad y = x - r \rightarrow x = y + r \quad Y = X - r$

If $x = r \quad y = 0$

$x = r+1 \quad y = 1$

$\vdots$

$$\binom{x-1}{r-1} p^r (1-p)^{x-r} I_{\{r, r+1, \ldots\}}(x)$$

$$\downarrow$$

$$\binom{r+y-1}{y} p^r (1-p)^y I_{\{0,1,\ldots\}}(y)$$

r+y trial success
need (r−1) fail
(r+y−1) trials
y → success
$(r+y-1) - (r-1) = y$

Sometimes see $\binom{r+y-1}{y} = (-1)^y \binom{-r}{y}$

$$\binom{r+y-1}{y} = \frac{(r+y-1)(r+y-2)\ldots(r)\overbrace{\cancel{(r-1)(r-2)\cdots(1)}}^{(r-1)!}}{y!\,[\cancel{(r-1)!}]}$$

$$= (-1)^y \frac{(-r)(-r-1)(-r-2)\ldots(-r-y+1)}{(y)(y-1)(y-2)\cdots(2)(1)} = (-1)^y \binom{-r}{y}$$

$$P(Y=y) = \underbrace{(-1)^y \binom{-r}{y} p^r (1-p)^y * I_{\{0,1,\ldots\}}(y)}$$

name: negative binomial
"striking resemblance".

$$E[Y] = \sum_{y=1}^{\infty} y \binom{r+y-1}{y} p^r (1-p)^y$$

$$= \sum_{y=1}^{\infty} \frac{\cancel{y}\,(r+y-1)!}{\cancel{y}\,(y-1)!\,(r-1)!} p^r (1-p)^y = \sum_{y=1}^{\infty} \frac{r\,(r+y-1)!}{r\,(y-1)!\,(r-1)!} = \sum_{y=1}^{\infty} r \binom{r+y-1}{y-1} p^r (1-p)^y$$

$$z = y-1 \Rightarrow \sum_{y=0}^{\infty} r \binom{r+z}{z} p^{r-1+1} (1-p)^{z+1} = r \frac{(1-p)}{p} \underbrace{\sum_{z=0}^{\infty} \binom{(r+1)+z-1}{z} p^{r+1} (1-p)^z}_{\text{negative binomial pmf}}$$

$$E[Y] = \frac{r(1-p)}{p}$$

Similarly... $Var(Y) = \dfrac{r(1-p)}{p^2}$

Uses in medical research: Define $\mu = \dfrac{r(1-p)}{p}$

$Var(Y) = \mu + \frac{1}{r}\mu^2$

Variance is quadratic ft'n of mean

Analyze count data (vs Poisson) Poisson limiting case
$r \to \infty$
$p \to 1 \quad \ldots$

Geometric Dist'n (special case of neg binom dist'n, $r=1$)

trials to 1st success ($r=1$)

$$P(X=x|p) = p(1-p)^{x-1} \quad x = 1, 2, \ldots$$

Show pdf sums to 1 (recall $|a|<1$: $\sum_{x=1}^{\infty} a^{x-1} = \frac{1}{1-a}$)

$$p\sum_{x=1}^{\infty}(1-p)^{x-1} = p\sum_{y=0}^{\infty}(1-p)^{y} = \frac{p}{1-(1-p)} = 1$$

$E[X] = \frac{1}{p}$ $\quad Var(X) = \frac{1-p}{p^2}$ $\Big\rangle$ from neg binom.

Cool Property: 'memoryless'

$$P(X>s|X>t) = P(X>s-t)$$

If $s=5$, $t=2$

$$P(X>s|X>t)$$

$$= \frac{P(X>s \text{ and } X>t)}{P(X>t)}$$

$$= \frac{P(X>s)}{P(X>t)}$$

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

$$= (1-p)^{s-t}$$

$$= P(X>s-t)$$

$$P(X>5|X>2) = P(X>3)$$

$$\frac{P(X>5 \text{ and } X>2)}{P(X>2)}$$

$$= \frac{P(X>5)}{P(X>2)} = \frac{(1-p)^5}{(1-p)^2} = (1-p)^3$$

Aside $P(X>s)$

$= P(0 \text{ success } s \text{ trials})$

$= (1-p)^s$

wikipedia (accessed 9/24/18)

Gottfried Wilhelm Leibnitz
German 1646 - 1716

- Conceiving ideas of differential and integral calculus independently of Newton's (contemporaneous).

  Leibnitz calc notation preferred over Newton's.

- refined binary number system

James Bernoulli 1655 - 1705
Swiss

- Discoverer of 'e'.
- Derived 1st version of 'Law of Large Numbers'
- Sided w/ Leibnitz during Leibniz - Newton calculus controversy.

Baron Siméon Denis Poisson 1781 - 1840
French

- contributions to math, physics & engineering
- advisor to Dirichlet