

Stats 117 Lecture 12

Named Distributions

Prof. John Duchi

Autumn 2026

12-1

Organization

- ▶ recap: Bernoulli and binomial
- ▶ sampling without replacement: hypergeometric
 - ▶ the lady tasting tea
- ▶ rare events: Poisson
- ▶ (time permitting) waiting for r successes: negative binomial

12-2

Recap: Bernoulli and binomial

- ▶ $X \sim \text{Bin}(n, p)$ counts heads in n independent tosses of a coin with $P(\text{heads}) = p$:

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

- ▶ $\text{Bernoulli}(p) = \text{Bin}(1, p)$
- ▶ if $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$, then $X_1 + \dots + X_n \sim \text{Bin}(n, p)$
- ▶ mean and variance:

$$\mathbf{E}[X] = np, \quad \text{Var}(X) = np(1-p)$$

12-3

Quality control: sampling with replacement

lot of $N = 2000$ cans, $M = 150$ defective; inspect $n = 60$ cans (with replacement), reject lot if 5 or more defective

- ▶ $X = \text{number of defectives in sample} \sim$

- ▶ $P(\text{reject}) = P(X \geq 5) =$

- ▶ $P(X \geq 5) \approx .4716, \quad \mathbf{E}[X] = n \frac{M}{N} = 4.5$

12-4

Sampling without replacement

don't inspect same can twice

- ▶ *simple random sample*: draw n of N units without replacement
- ▶ draws no longer independent

counting:

- ▶ number of possible samples:

- ▶ number of samples with exactly k defectives:

12-5

Hypergeometric distribution

Definition

$X \sim \text{HyperGeom}(n, M, N)$ counts defectives in simple random sample of size n from N units, M of them defective:

$$p_X(k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \dots, \min\{n, M\}$$

12-6

Quality control: without replacement

same lot: $N = 2000$, $M = 150$, $n = 60$

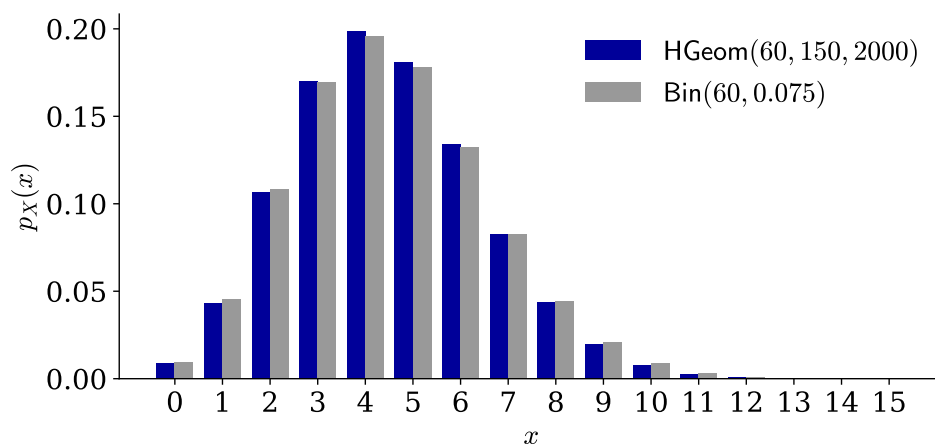
► $X \sim$

► $P(X \geq 5) =$

► $P(X \geq 5) \approx .4729$ (with replacement: .4716)

12-7

Hypergeometric versus binomial



► hyper has less mass in upper tail

12-8

Mean and variance of the hypergeometric

Proposition

If $X \sim \text{HyperGeom}(n, M, N)$ and $p = \frac{M}{N}$, then

$$\mathbf{E}[X] = n \frac{M}{N} = np, \quad \text{Var}(X) = np(1-p) \cdot \frac{N-n}{N-1}.$$

► *finite population correction* $\frac{N-n}{N-1}$

► cans: $\frac{2000-60}{2000-1} \approx .97$

what happens as $N \rightarrow \infty$?

12-9

The lady tasting tea

Muriel Bristol claims she can tell whether milk was poured before or after the tea; R. A. Fisher tests her

- 8 cups: 4 milk-first, 4 tea-first, in random order
- she must pick the 4 milk-first cups; she gets all 4 right

if she is just guessing:

► $X = \text{number correct} \sim$

► $P(X = 4) =$

12-10

The lady tasting tea: culture

- ▶ *p-value* $P(X = 4) = \frac{1}{70} \approx .014$
- ▶ small, so Fisher rejects “just guessing” hypothesis
- ▶ now called *Fisher’s exact test*

12-11

Puzzle: pots of coins

n people each put coin in a pot; each coin goes to one of the n people uniformly at random, independently

- ▶ as $n \rightarrow \infty$, what is $P(\text{you get nothing})$?

two wrong arguments:

1. pot becomes infinite, so probability $\rightarrow 0$
2. chance $\frac{1}{n}$ of each coin $\rightarrow 0$, so probability $\rightarrow 1$

12-12

The pot o' coins: solution

X = coins you receive $\sim$

$$P(X = 0) =$$

n	2	5	10	100	1000	∞
$(1 - \frac{1}{n})^n$	.25	.328	.349	.366	.368	$e^{-1} \approx .3679$

12-13

Poisson approximation to the binomial

Theorem

Let $X \sim \text{Bin}(n, \mu/n)$ for a fixed $\mu > 0$. As $n \rightarrow \infty$,

$$p_X(k) \rightarrow e^{-\mu} \frac{\mu^k}{k!}, \quad k = 0, 1, 2, \dots$$

proof:

$$p_X(k) = \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k}$$

12-14

Poisson approximation: proof continued

$$p_X(k) = \underbrace{\frac{\mu^k}{k!}}_{(a)} \cdot \underbrace{\frac{n!}{(n-k)! n^k}}_{(b)} \cdot \underbrace{\left(1 - \frac{\mu}{n}\right)^n}_{(c)} \cdot \underbrace{\left(1 - \frac{\mu}{n}\right)^{-k}}_{(d)}$$

12-15

Poisson distribution

Definition

$X \sim \text{Poisson}(\mu)$ for $\mu > 0$ if

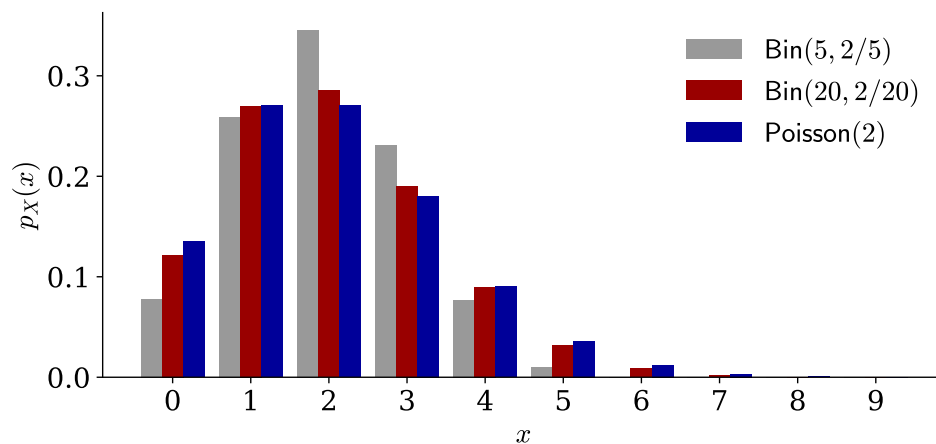
$$p_X(k) = e^{-\mu} \frac{\mu^k}{k!}, \quad k = 0, 1, 2, \dots$$

► check: $\sum_{k=0}^{\infty} p_X(k) = 1$

► counts of rare events: many chances, each unlikely

12-16

Binomial versus Poisson



- $\text{Bin}(n, \mu/n)$ approaches $\text{Poisson}(\mu)$ as n grows

12-17

Mean and variance of the Poisson

Proposition

If $X \sim \text{Poisson}(\mu)$, then $\mathbf{E}[X] = \mu$ and $\text{Var}(X) = \mu$.

guess: $np = n \cdot \frac{\mu}{n} = \mu$ (not a proof but eh, good enough)

12-18

Detecting radiation

Geiger counter clicks on average 45 times / minute

► X = number of clicks in one second

► $X \sim \text{Poisson}(.75)$

► $P(X > 1) \approx .17336$

12-19

Negative binomial distribution

Definition

$X \sim \text{NegBin}(r, p)$ counts tosses until r heads:

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \dots$$

why: (where must the last head be?)

► $r = 1$:

12-20

Mean and variance of the negative binomial

Proposition

If $X \sim \text{NegBin}(r, p)$, then

$$\mathbf{E}[X] = \frac{r}{p}, \quad \text{Var}(X) = \frac{r(1-p)}{p^2}.$$

► $r = 1$: geometric has $\mathbf{E}[N] = \frac{1}{p}$ and $\text{Var}(N) = \frac{1-p}{p^2}$

intuition: (each head takes about $1/p$ tosses)

12-21

Recruiting for a clinical trial

need 20 subjects; each interviewee eligible with probability .6, independently

► X = number of interviews $\sim$

► $\mathbf{E}[X] =$

► $P(X \geq 40) =$

► $P(X \geq 40) \approx .1021$

12-22