5/28/19 Lecture 17

Def: If the X_i in $X_1, X_2, \ldots$ are IID Bernoulli(p), then $(X_1, X_2, \ldots)$ are called Bernoulli trials with parameter p; if the sequence $(X_1, X_2, \ldots)$ is infinite

This defines a Bernoulli (stochastic) process

Binomial

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

(ie X follows the Binomial distribution with parameters n (positive integer) and $0 < p < 1$

$$\leftrightarrow f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \underbrace{I\{0, 1, \ldots, n\}(x)}_{\text{support}(X)}$$

Consequence

1) $X_1, \ldots, X_n \overset{IID}{\sim} \text{Bernoulli}(p)$

$$\rightarrow X = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$$

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

$E(X) = np$

$V(X) = np(1-p)$

$\psi_X(t) = [pe^t + (1-p)]^n$ for all $-\infty < t < \infty$

$SD[X] = \sqrt{np(1-p)}$

Case Study

Supreme Court Case: Casteneda v. Partida (1977)

Grand juries in the U.S. judicial system have catchment areas: everybody 18 and over living in the judicial district for that grand jury (& a few other minor restrictions)

Hidalgo county, Texas
eligible pool was 79.1% Mexican-American
2½ year period at issue in Supreme Court Case:
220 people called to serve on grand juries, but only 100 of them were Mexican-American

Q: Prima facie case of discrimination?

Before this 2½ year period, let $\overline{X}$ be your prediction of # of Mexican-Americans among the 220 people

If no discrimination, $\overline{X} \sim \text{Binomial}(220, 0.791)$
$(\overline{X} \mid T_1)$ ↗

T_1 = theory 1 = no discrimination

$$E(\overline{X} \mid T_1) = \underset{n}{(220)} \cdot \underset{p}{(0.791)} \doteq 174.0$$

$$SD(\overline{X} \mid T_1) = \sqrt{np(1-p)} \doteq 6.0$$

Q: If you were expecting 174 give or take 6, would you be surprised to see 100?

A: You'd be astonished

Frequentist statistical answer:

$P(\bar{X} \leq 100 \mid T_1) = 8.0 \times 10^{-28}$ so T_1 looks ridiculous

Bayesian statistical answer:

Need to compute $P(T_1 \mid \bar{X} = 100)$, not the other way around

Hypergeometric

A finite population has A elements of type 1 and B elements of type 2

Total population size: (A+B)

You choose n elements at random w/out replacement from this population (ie you take a simple random sample (SRS) of size n)

Let $\bar{X}$ = # elements of type 1 in your sample

Then $\bar{X}$ follows the hypergeometric distribution with parameters (A, B, n)

The pmf of X is

$$f_X(x \mid A, B, n) = \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}} \; I\left[\max\{0, n-B\} \le x \le \min\{n, A\}\right]$$

for (A, B, n) non-negative integers with $n \le A+B$

Consequences

1) $E(X) = n \cdot \frac{A}{A+B}$

2) $V(X) = n \cdot \left(\frac{A}{A+B}\right)\left(\frac{B}{A+B}\right)\left(\frac{A+B-n}{A+B-1}\right)$

Note that if your sampling had been with replacement (ie you take an IID sample), X would have been Binomial with the same value of n and $p = \frac{A}{A+B}$

In that case $E(X) = np = n\frac{A}{A+B}$ and

$V(X) = np(1-p) = n\left(\frac{A}{A+B}\right)\left(\frac{B}{A+B}\right)$ (compare)

If you let $T = (A+B)$ be the total # of elements in the population, ...

sampling method	mean	variance
with repl. (IID)	$n\left(\frac{A}{A+B}\right)$	$n\left(\frac{A}{A+B}\right)\left(\frac{B}{A+B}\right)$
without repl. (SRS)	$n\left(\frac{A}{A+B}\right)$	$n\left(\frac{A}{A+B}\right)\left(\frac{B}{A+B}\right)\left(\frac{T-n}{T-1}\right)$

$0 \leq d = \frac{T-n}{T-1} \leq 1$ is called the finite population correction

3 special cases worth considering

(a) $(n=1)$ $d=1 \leftrightarrow$ SRS = IID, with only 1 element sampled

(b) $(n=T)$ $d=0 \leftrightarrow$ If you exhaust the entire pop. with SRS, you have no uncertainty left

(c) (n fixed, $T \uparrow$) $d \uparrow 1 \leftrightarrow$ With a small sample from a large population, SRS = IID

Poisson

$(\lambda > 0)$ $X \sim \text{Poisson}(\lambda) \leftrightarrow X$ has pmf:

$$f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} \underbrace{I\{0,1,\ldots\}(x)}_{\text{support of } X}$$

$E(X) = \lambda$ $V(X) = \lambda$ Thus for the Poisson dist $\frac{V(X)}{E(X)} = 1$

Def: If $E(X)$ and $V(X)$ both exist and $E(X) \neq 0$

$\frac{V(X)}{E(X)}$ is called the variance-to-mean ratio (VTMR)

$$\psi_X(t) = e^{\lambda(e^t - 1)} \qquad -\infty < t < \infty$$

The Poisson can be unrealistic as a consequence of its VTMR of 1, many rvs that represent counts of occurrences of events in time intervals of fixed length have VTMR > 1

The Poisson and Binomial distributions both count the number of "successes" in a process unfolding in time, so it should not be surprising to find out that these 2 dist. are related:

When (n is large, p is close to 0), Binomial$(n, p) \doteq$ Poisson$(n \cdot p)$

Thm: n positive integer, $0 < p < 1$
$X \sim$ Binomial(n, p)

$\lambda > 0$, $Y \sim$ Poisson(λ)

Choose any sequence $\{p_n\}_{n=1}^{\infty}$ of values between 0 and 1 with $\lim_{n \to \infty} n \cdot p_n = \lambda$

Then $f_X(x \mid n, p_n) \xrightarrow[n \to \infty]{} f_Y(y \mid \lambda)$

Poisson process revisited

Def: A Poisson process with rate λ per unit time (or space, volume, etc) is a stochastic process with two properties:

(a) # arrivals in every interval of time of length $t \sim$ Poisson(λt)

(b) #s of arrivals in all disjoint (non-overlapping) time intervals are independent

Case Study

Parasitic Protozoa in drinking water

There's a kind of parasitic organism called cryptosporidium that's capable of getting into the public drinking water supplies; at one stage in their life cycle they're called oocysts.

They can make people sick at a concentration of only 1 oocyst per 5 liters = 1.3 gallons of water

One problem is that it can be hard to detect these oocysts with water filtration

Suppose that in the water supply of your city, oocysts occur according to a Poisson process with rate λ oocysts per liter, and that the filtering system your water utility company uses can capture all the oocysts in a water sample but only has probability p of detecting each oocyst that's actually there

Set Y = actual # oocysts in t liters of water and

$X_i = \begin{cases} 1 & \text{if oocyst } i \text{ gets counted} \\ 0 & \text{else} \end{cases}$

(and counting events are independent

X = # counted oocysts

Then $(X \mid Y = y) = \sum_{i=1}^{y} X_i$

Under these assumptions, $(X \mid Y = y) \sim \text{Binomial}(y, p)$

Q: What's the dist. of X?

A: By the Law of total probability

$$f_X(x) = P(X = x) = \sum_{y=0}^{\infty} P(Y = y) P(X = x \mid Y = y)$$

for all $x = 0, 1, \ldots$ in which $P(Y = y) = \frac{(\lambda t)^y e^{-\lambda t}}{y!}$

for $y = 0, 1, \ldots$ and $P(X = x \mid Y = y) = \binom{y}{x} p^x (1-p)^{y-x}$

Notice that if $X = x$, $Y \geq x$ because the actual number of oocysts (Y) has to be at least as large as the # of oocysts deleted (X)

After a careful calculation

$$f_X(x) = \sum_{y=x}^{\infty} \binom{y}{x} p^x (1-p)^{y-x} \frac{(\lambda t)^y e^{-\lambda t}}{y!}$$

you get $\frac{e^{-p\lambda t}(p\lambda t)^x}{\lambda!}$

$X \sim \text{Poisson}(p\lambda t)$: losing a proportion $(1-p)$ of the oocysts to faulty counting just lowers the rate of the Poisson process from λ/liter to $\lambda \cdot p$/liter (makes excellent sense)

In practice oocysts are hard to detect: p is small (not far from 0)

Q: How much water (t liters) do you need to filter to achieve $P(\text{at least 1 oocyst detected}) \geq 1-\alpha$ for small α?

A: Not hard to work out

$P(\text{at least 1 detected}) = 1 - P(\text{none detected})$

$= 1 - P(X = 0) = 1 - e^{-p\lambda t} \geq 1 - \alpha$

$\leftrightarrow \alpha \geq e^{-p\lambda t} \leftrightarrow \ln \alpha \geq -p\lambda t \leftrightarrow t \geq \frac{\ln \alpha}{p\lambda}$

Ex: $\alpha = .01$, $p = 0.1$

$\lambda = 0.2$ / liter (1 per 5 liters) ← minimum sickness level

To achieve prob 99% + has to be at least 230.3 liters

Negative Binomial Distribution

You're watching a potentially endless sequence of Bernoulli trials with constant success probability p.

Let X = # failures before rth success

rth integer ≥ 1

The pmf of X that follows the Negative Binomial dist:

$$f_X(x \mid r, p) = \binom{r+x-1}{x} p^r (1-p)^x \mathbb{I}_{\{0,1,\ldots\}}(x)$$

with parameters (r, p) $(0 < p < 1)$

The name comes from the fact that, when you watch a sequence of Bernoulli trials with constant unknown success probability p unfold, there are two different ways to estimate p: decide ahead of time to sample n (known constant) success/failure trials and record the (random) # S of successes you see

(from which a reasonable estimate would be $\hat{p}_B = \frac{S}{n}$ ← binomial)

or decide ahead of time that you're going to sample until you've seen S (known constant) successes & record the (random) # of trials N needed to accumulate that many successes (from which a reasonable estimate would be $\hat{p}_{NB} = \frac{S}{N}$ ← Negative binomial)

Special Case of Negative Binomial

Set $r=1$ and record the number X of failures until the first success: X is said to follow the Geometric(p) distribution, with pmf

$$f_X(x|p) = p(1-p)^x \underbrace{I_{\{0,1,\ldots\}}(x)}_{\text{support } X}$$

(parameter p)

Consequence

$X_1, \ldots, X_n$ IID Geometric(p) $\rightarrow \sum_{i=1}^{n} X_i \sim$ Negative Binomial(n, p)

This is a direct analogue to the Bernoulli/Binomial story: $X_1, \ldots, X_n$ IID

Bernoulli$(p) \rightarrow \sum_{i=1}^{n} X_i \sim$ Binomial(n,p)

$X \sim$ Negative Binomial (n,p)

$$\psi_X(t) = \left[\frac{p}{1-(1-p)e^t}\right]^r \text{ for } t < \log\left(\frac{1}{1-p}\right)$$

from which $E(X) = \frac{r(1-p)}{p}$, $V(X) = \frac{r(1-p)}{p^2}$

Consequence

$X \sim$ Geometric$(p) \rightarrow P(X = k+t \mid X \geq k) = P(X = t)$

$\{k\}$ $\{t\}$ both non-negative integers

This is called the memoryless property of the Geometric distribution, and it turns out that this is the only discrete distribution with this property

X = # failures until first success = 5 (here)

F F F F F S

X

Y

Y = # failures, starting at trial ($k+1 = 2$ here) until next success ($= 4$ here)

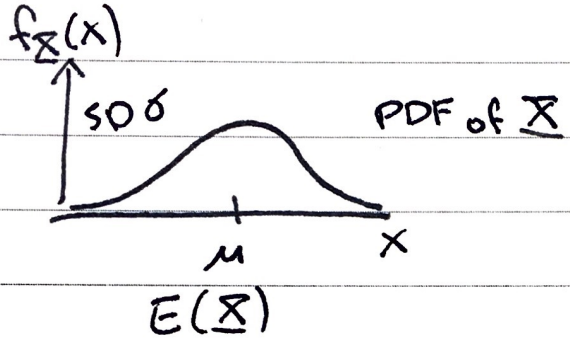
Then Y has the same dist. as X and is independent of what happened on the first k trials ie "The process has no memory."

Case 2: Important continuous distributions

Normal (Gaussian) Distribution

$X \sim \text{Normal}(\mu, \sigma^2)$ mean μ and variance $0 < \sigma^2 < \infty$

PDF $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$



The Normal dist. is the single most important dist. in all of probability and statistics (mainly for 2 reasons):

1) Many observable random processes have dist. shapes that are close to the "bell curve" (Normal PDF)

2) The Central Limit Theorem (CLT)

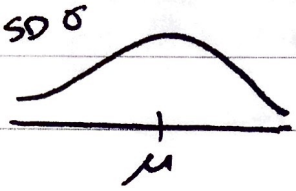
Properties of the Normal Dist.

$X \sim \text{Normal}(\mu, \sigma^2)$

$E(X) = \mu$

$V(X) = \sigma^2$ $\quad \Psi_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$

$SD(X) = \sigma$



(Center of symmetry, mean, median, mode) = all μ

Consequences

1) $X \sim \text{Normal}(\mu, \sigma^2)$

$Y = aX + b$ $\begin{pmatrix} a \neq 0 \\ b \end{pmatrix}$ fixed constants

$Y \sim \text{Normal}(a\mu + b, a^2\sigma^2)$

In other words, Normality is preserved under linear transformation

Def: The Normal dist. with mean $\mu = 0$ and SD $\sigma = 1$ is the standard normal dist

The PDF of $X \sim \text{Normal}(0, 1)$ is $\phi_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ and its CDF is $\Phi(x) \triangleq \int_{-\infty}^{x} \phi_X(t)\, dt$

Empirical Rule

Part 1: Start at the mean μ of a distribution and go 1 SD σ either way: you will find (about 2/3) 68% of the probability in the interval $(\mu \pm 1\sigma)$

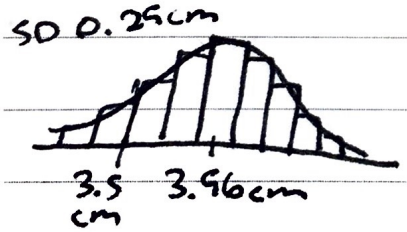
Part 2: Ditto 2 SDs either way: $(\mu \pm 2\sigma)$ captures most (about 95%) of the probability

Part 3: Ditto 3 SDs either way: $(\mu \pm 3\sigma)$ captures almost all (99.7%) of the probability

This rule is exact for all Normal dists. It is surprisingly good approximation for many other distributions

This permits an easy trick that's helpful in computing Normal probabilities

Ex: You have a random sample of $n=103$ immature monarch butterflies, and you measure their wing lengths:



histogram of wing length

y = ring length (cm)

$$\begin{bmatrix} y_1 = 4.1 \\ y_2 = 3.3 \\ \vdots \\ y_n = 4.7 \end{bmatrix} \quad n=103$$

mean $\bar{y} = 3.96$cm

SD $s = 0.29$cm

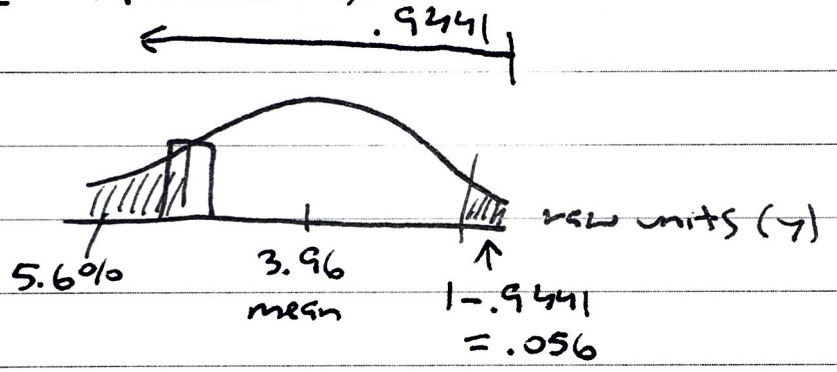
Q: About what % of the sampled butterflies had wing length ≤ 3.5 cm?

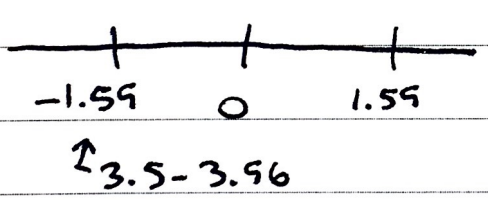
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

↑ sample mean ↑ sample SD

A_1 (exact) $\frac{8}{103} \doteq 7.8\%$

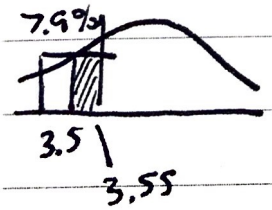
A_2 (approximate)





Converting to su for data: $z = \frac{y - \bar{y}}{s} = su$

for random variables $z = \frac{Y - \mu}{\sigma} = su$



Keeping track of histogram bar edges: continuity correction

More consequences

4) $\bar{X}_1, \ldots, \bar{X}_k$ independent, $\bar{X}_i \sim \text{Normal}(\mu_i, \sigma_i^2)$

$$\rightarrow \sum_{i=1}^{k} X_i \sim \text{Normal}\left(\sum_{i=1}^{k} \mu_i, \sum_{i=1}^{k} \sigma_i^2\right)$$

This additive property is why Normal dists. are indexed by variance rather than SD

Notation

$\text{Normal}(\mu, \sigma^2) \triangleq N(\mu, \sigma^2)$

Ex: Population of adult U.S. women: height follows $N(\mu = 65.0\,in, \sigma^2 = 3.2^2\,in^2)$ dist
$(\sigma = 3.2\,in)$

Pop. of adult U.S. men: height follows $N(\mu = 69.5\,in, \sigma^2 = 3.3^2\,in^2)$ dist

1 woman chosen at random: height W

1 man chosen at random (independently): height M

$P(\text{woman taller than man}) = P(W > M) = ?$

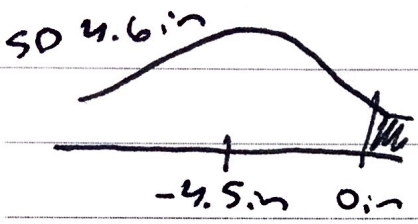
Define $D = W - M$

By consequence ⑨, $D \sim N(65 - 69.5 = -4.5\,in, 3.2^2 + 3.3^2 = 21.1\,in^2)$

$P(W > M) = P(D > 0)$

$SD\sqrt{21.1\,in^2} = 4.6\,in$

Convert to S_w: $\frac{0-(-4.5)}{4.6} = 0.98$



PDF of D

New units

So $P(w > M) = 16\%$

(about 1 in 6)