Def If the X_i in $X_1, X_2, \ldots$ are IID Bernoulli (p), then $(X_1, X_2, \ldots)$ are called <u>Bernoulli trials</u> with <u>parameter</u> p; if the sequence $(X_1, X_2, \ldots)$ is infinite this defines a <u>Bernoulli (stochastic) process</u>.

Binomial $X \sim$ Binomial(n, p) (i.e., X follows the Binomial distribution with parameters n $\left(\begin{array}{c}\text{positive}\\ \text{integer}\end{array}\right)$ and $0 < p < 1$)

$$\leftrightarrow \quad f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \overbrace{I_{\{0, 1, \ldots, n\}}}^{\text{support}(X)}(x)$$

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

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

$X \sim \text{Binomial}(n, p)$ | $E(X) = n \cdot p$ | $V(X) = np(1-p)$

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

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

Case Study | Supreme Court case: Castaneda v. Partida (1977)

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

Hidalgo County, Texas — extreme southern border of TX with Mexico

eligible pool was 79.1% Mexican-American

$2\frac{1}{2}$ yr 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 $2\frac{1}{2}$ yr period, let X be your prediction of # of Mexican-Americans among the 220 people

If no discrimination,

$X \sim \text{Binomial}(220, 0.791)$
$(X \mid T_1)$

T_1 = theory 1 = no discrimination

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

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

$\boxed{Q:}$ If you were expecting 174 give or take 6, would you be surprised to see 100?

$\boxed{A:}$ You'd be astonished

Frequentist statistical answer

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

Bayesian statistical answer

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

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 without replacement from this population (i.e., you take a simple random sample (SRS) of size n) | Let X = (# elements of type 1 in your sample)

Then (as noted in take-Home Test 1 problem 2) X follows the hypergeometric distribution with parameters (A, B, n). As we saw in that problem, the PMF of X is

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

support(X)

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

Consequences

① $E(X) = n \cdot \frac{A}{A+B}$

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

Note that if your sampling had been with replacement (i.e., 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)	$n\left(\frac{A}{A+B}\right)$	$n\left(\frac{A}{A+B}\right)\left(\frac{B}{A+B}\right)\left(\frac{T-n}{T-1}\right)$

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

3 special cases worth considering

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

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

(c) (n fixed, $T \uparrow$) $\alpha \uparrow 1$ $\leftrightarrow$ with a small sample from a large population,

SRS $\doteq$ IID

Poisson $(\lambda > 0)$ $\underline{X} \sim \text{Poisson}(\lambda)$

$\leftrightarrow$ X has PMF $f_X(x) = \dfrac{\lambda^x e^{-\lambda}}{x!} \underbrace{I_{\{0,1,\dots\}}(x)}_{\text{support of } X}$

$\boxed{E(X) = \lambda}$
$\boxed{V(X) = \lambda}$
} thus for the Poisson dist.

$\dfrac{V(X)}{E(X)} = 1$

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

$-\infty < t < \infty$

$\boxed{\text{Def.}}$ If $E(X)$ and $V(X)$ both exist and $E(X) \neq 0$, $\dfrac{V(X)}{E(X)}$ is called the variance-to-mean ratio (VTMR)

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

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

---

when $\left(\begin{array}{l} n \text{ is large} \\ p \text{ is close to } 0 \end{array} \right)$, $\text{Binomial}(n, p) \doteq \text{Poisson}(n \cdot p)$

Theorem n positive integer, $0 < p < 1$, $X \sim \text{Binomial}(n, p)$, $\lambda > 0$, $Y \sim \text{Poisson}(\lambda)$ / 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 $f_X(x \mid n, p_n) \underset{n \to \infty}{\longrightarrow} f_Y(y \mid \lambda)$

Poisson process, revisited

Def

A poisson process with rate λ per unit time (or space, or volume, or...) 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 249
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, & 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. (# counting events are independent)

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

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

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 $P(Y = y) = \dfrac{(\lambda t)^y e^{-\lambda t}}{y!}$ 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 number of oocysts detected (X). After a careful calculation you get:

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

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

i.e., $X \sim \text{Poisson}(p\lambda t)$: 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). $\boxed{Q:}$ How much water (t liters) do you need to filter to achieve $P(\text{at least 1 oocyte detected}) \geq 1-\alpha$ for small α? $\boxed{A:}$ Not hard to work out

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

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

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

$$\boxed{t \geq -\frac{\ln\alpha}{p\lambda}}$$

Example: $\alpha = .01$, $p = 0.1$, $\lambda = 0.2/\text{liter}$ (1 per 5 liters)
$\updownarrow$
(minimum sickness level)

to achieve prob 99%, t has to be at least $\underline{230.3}$ liters.

Negative Binomial Distribution

You're watching a potentially endless sequence of Bernoulli trials with constant success probability p. Let X = # failures before r^{th} success (r integer ≥ 1)

You can show that X follows what's called the Negative Binomial dist:

its PMF is $f_X(x \mid r, p) = \binom{r+x-1}{x} p^r (1-p)^x \, I_{\{0,1,2,\ldots\}}(x)$.

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 of X failures until the first success: X is said to follow the

Geometric (p) distribution, with

PMF (parameter p)

$$f_X(x \mid p) = p(1-p)^x \, I_{\{0,1,\ldots\}}(x).$$

support of X

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).

$\boxed{X \sim \text{Negative Binomial } (r, p)}$

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

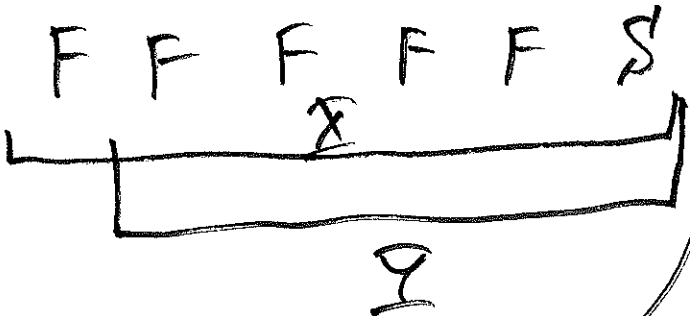
from which $E(X) = \dfrac{r(1-p)}{p}$, $V(X) = \dfrac{r(1-p)}{p^2}$

Consequence $X \sim \text{Geometric}(p) \rightarrow$

$\left\{\begin{matrix} k \\ t \end{matrix}\right\}$ both non-negative integers

$$P(X = k+t \mid X \geq k) = P(X = t)$$

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 <u>first</u> success $= 5$ (here)

Y = # failures, starting at trial $\begin{pmatrix} k+1 \\ =2 \end{pmatrix}$ here, until <u>next</u> success ($= 4$ here)

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

Case 2: Important Continuous Distributions

Normal (Gaussian) Distribution

$X \sim \text{Normal}(\mu, \sigma^2)$ mean μ, 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 & statistics, mainly for 2 reasons: ① many observable random processes have dist. shapes that are close to the "bell curve" (Normal PDF), and ② the Central Limit Theorem (CLT), which we'll examine soon.

Properties of the Normal Dist.

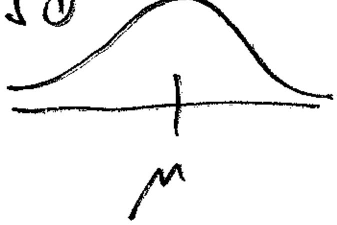
$X \sim \text{Normal}(\mu, \sigma^2)$ $N(\mu, \sigma^2)$

$E(X) = \mu$

$V(X) = \sigma^2$, $SD(X) = \sigma$

$$\psi_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

SD σ



dist. of X

center of symmetry, mean, median, mode — all $= \mu$

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

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

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

In other words, Normality is preserved under linear transformation | Def.

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

The PDF of $X \sim$ Normal $(0, 1)$ is

$$\phi_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$ and its

phi (lower-case)

CDF is $\Phi(x) \triangleq \int_{-\infty}^{x} \phi_X(t)\, dt$

uppercase Phi Φ

It turns out that e^{-cx^2} $(c>0)$ has no

anti-derivative in closed form, so

$\Phi(x)$ cannot be summarized in a

formula; instead it's approximated by

numerical integration (see p. 861 in DS).

Consequences, continued

② Because the Normal PDF (for all $x \in \mathbb{R}$) is symmetric, $\Phi(-x) = 1 - \Phi(x)$

and $\Phi^{-1}(p) = -\Phi^{-1}(1-p)$ (for all $0 < p < 1$)

---

③ $X \sim \text{Normal}(\mu, \sigma^2) \rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$,

so that $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

and $F_X^{-1}(p) = \mu + \sigma\,\Phi^{-1}(p)$

Empirical Rule

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

Part 2 Ditto 2 SDs either way: $(\mu \pm 2\sigma)$ captures a (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 & is a surprisingly good approximation for many other distributions. This permits an easy trick that's helpful in computing Normal probabilities.

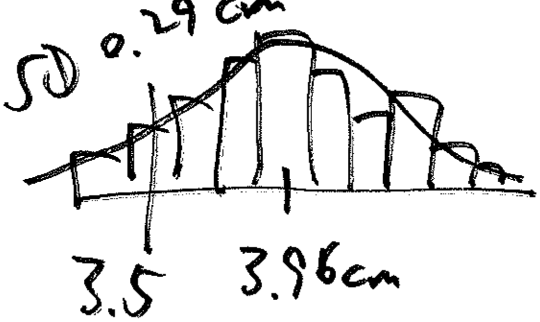
Example: You have a random sample of $n = 103$ immature monarch butterflies, and you measure their wing lengths:

y = wing 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?

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

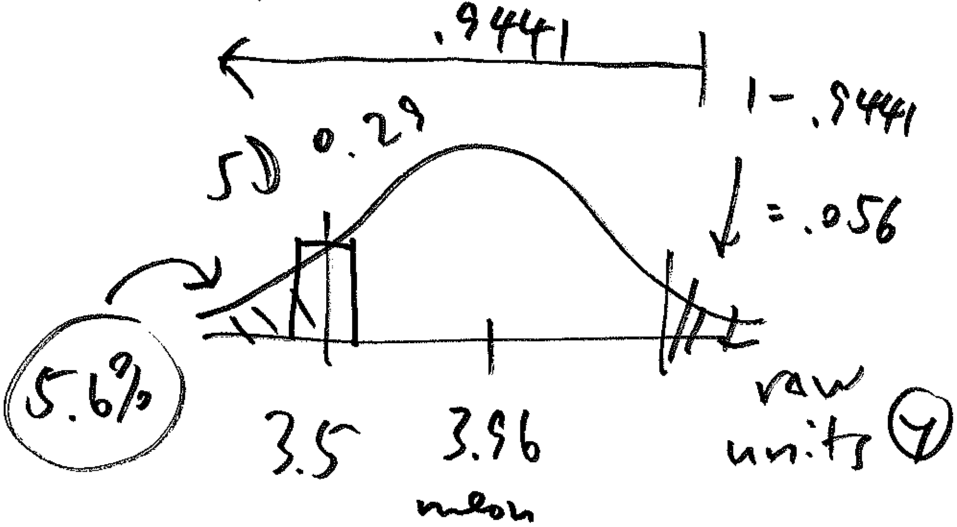
sample mean sample SD

sorted y

3.2	4	103
3.3		
⋮		
3.5	8	
3.5		
3.5		
3.5		
3.6		
⋮		
4.7		

$\boxed{A_1}$ (exact) $\frac{8}{103} \doteq 7.8\%$ (263)

$\boxed{A_2}$ (approximate)



−1.59	0	+1.59	standard units (z) (SU)

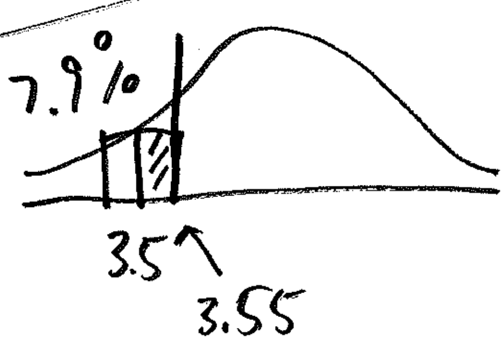
$$\frac{3.5 - 3.96}{0.29}$$

converting to SU for data:

$$z = \frac{y - \bar{y}}{s} = SU$$

for random variables

$$Z_Y = \frac{Y - \mu}{\sigma} = SU$$



keeping track of histogram bar edges: continuity correction

More consequences

(4) $X_1, \ldots, X_k$ independent, $X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$

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

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

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

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

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

1 woman chosen at random, height $\underline{W}$; (265)

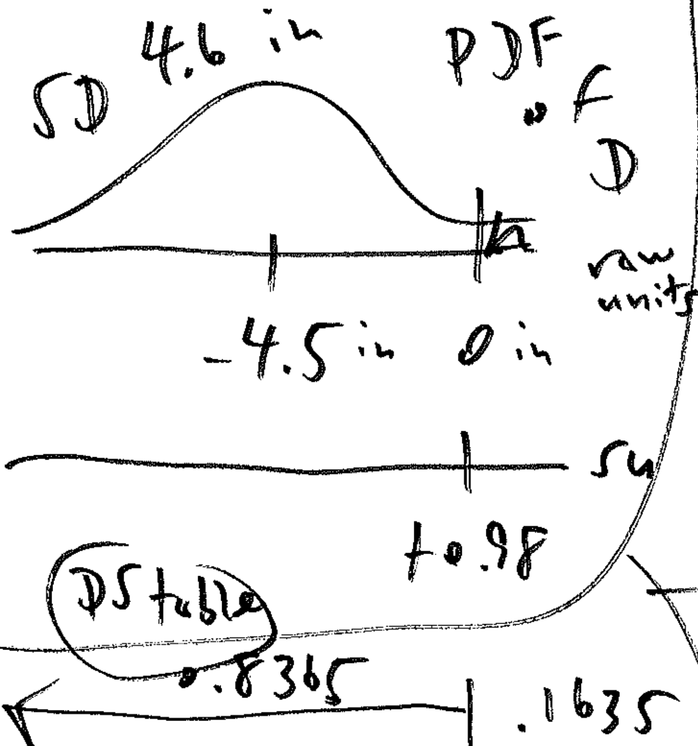
1 man chosen at random (independently), height $\underline{M}$; P(woman taller than man) $= P(\underline{W} > \underline{M})$ = ?

Define $\underline{D} = \underline{W} - \underline{M}$

By consequence (4), $\underline{D} \sim N(65 - 69.5 = -4.5$ in, $3.2^2 + 3.3^2 = 21.1$ in$^2)$

SD $\sqrt{21.1 \text{ in}^2} \doteq 4.6$ in

$P(\underline{W} > \underline{M}) = P(\underline{D} > 0)$



Convert to su:

$$\frac{0 - (-4.5)}{4.6} \doteq +0.98$$

so $P(\underline{W} > \underline{M}) \doteq 16\%$ (about 1 in 6)

(28 May 19)