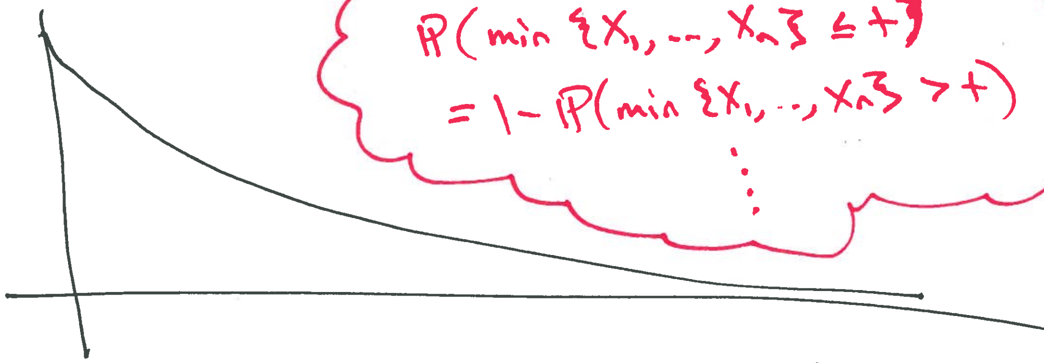
EX: Suppose that the circuits in some system have a lifetime that is exponentially distributed with parameter. λ.

$W = \min\{X_1, \dots, X_n\}$

$\mathbb{P}(\min\{X_1, \dots, X_n\} \leq t)$

$= 1 - \mathbb{P}(\min\{X_1, \dots, X_n\} > t)$

$\vdots$



Suppose that the system runs on n circuits and it operates until all n circuits die. Assume that the circuits are independent of one another.

Qn: What is the distribution (i.e., the CDF) of the lifetime of the system?

ANS: Let X_i be the lifetime of the i^{th} circuit. Then the random vector $(X_1, \dots, X_n)$ has all the information we need to determine the lifetime of the system.

We are defining a random variable $Z = \max\{X_1, \dots, X_n\}$.

we want $F_Z(t)$.

$$
\begin{aligned}
F_Z(t) &= \mathbb{P}(Z \leq t) \\
&= \mathbb{P}(\max\{X_1, \dots, X_n\} \leq t) \\
&= \mathbb{P}(X_1 \leq t, X_2 \leq t, \dots, X_n \leq t) && \text{(by definition of max)} \\
&= \mathbb{P}(X_1 \leq t)\mathbb{P}(X_2 \leq t)\cdots\mathbb{P}(X_n \leq t) && \text{(by independence)} \\
&= (1-e^{-\lambda t})(1-e^{-\lambda t})\cdots(1-e^{-\lambda t}) && \text{(b/c each } X_i \sim \mathrm{Exp}(\lambda)) \\
&= (1-e^{-\lambda t})^n
\end{aligned}
$$

$\leftarrow$ CDF of the max, e.g., if you differentiated it w.r.t. t you would get the pdf

MAXIMUM LIKELIHOOD ESTIMATION

DEF: A statistic θ is just an function of a random sample.

DEF: A random sample is just a (typically finite) sequence of independent and identically distributed random variables.

EX: There are many possible statistics.

$$\theta_1(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n = \bar{X}$$

$$\theta_2(X_1, \ldots, X_n) = \max\{X_1, \ldots, X_n\}$$

$$\theta_3(X_1, \ldots, X_n) = \text{median}\{X_1, \ldots, X_n\}$$

$$\theta_4(X_1, \ldots, X_n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2$$

BIG PICTURE QUESTION: How do I find good statistics, i.e., statistics that are useful for estimating underlying population parameters?

DEF: the population parameters governing a r.v. are just those parameters that fully determine its distribution / CDF.

EX: $X \sim N(\mu, \sigma^2)$

$X \sim Sta(\alpha, \beta, \gamma, \delta)$ "stable r.v."

$X \sim Exp(\lambda)$

$X \sim Ber(p)$

DEF: Suppose that X_i is a r.v. with density $f_{X_i}(\theta; x_i)$. For a random sample $X_1, X_2, \ldots, X_n$ with common density $f_{X_i}(\theta; x_i)$, $i = 1, \ldots, n$, the likelihood function as

just semicolon → $\boxed{;}$

$$L(\theta; x_1, \ldots, x_n) = \overbrace{f_{X_1}(\theta; x_1)\, f_{X_2}(\theta; x_2) \cdots f_{X_n}(\theta; x_n)}^{\text{the joint density of } X_1, \ldots, X_n}$$

$$= \prod_{i=1}^{n} f_{X_i}(\theta; x_i)$$

$$= \prod_{i=1}^{n} f(\theta; x_i) \quad \text{(b/c identically distributed)}$$

$$\sum_{i=1}^{3} i^2 = 1^2 + 2^2 + 3^2$$

$$\prod_{i=1}^{3} (i+1) = 2 \cdot 3 \cdot 4$$

QN: Suppose that you have a random sample $X_1, \ldots, X_n$. You regard it as fixed. How do you choose a θ that best comports with, or is most compatible with, these data?

IDEA: Choose the θ that maximizes

$$L(\theta; \underbrace{X_1, \ldots, X_n}_{\text{realized values of our random sample}})$$

So, we will choose

$$\boxed{\hat{\theta} = \arg\max_{\theta} L(\theta; X_1, \ldots, X_n)}$$

↖ the proposal!

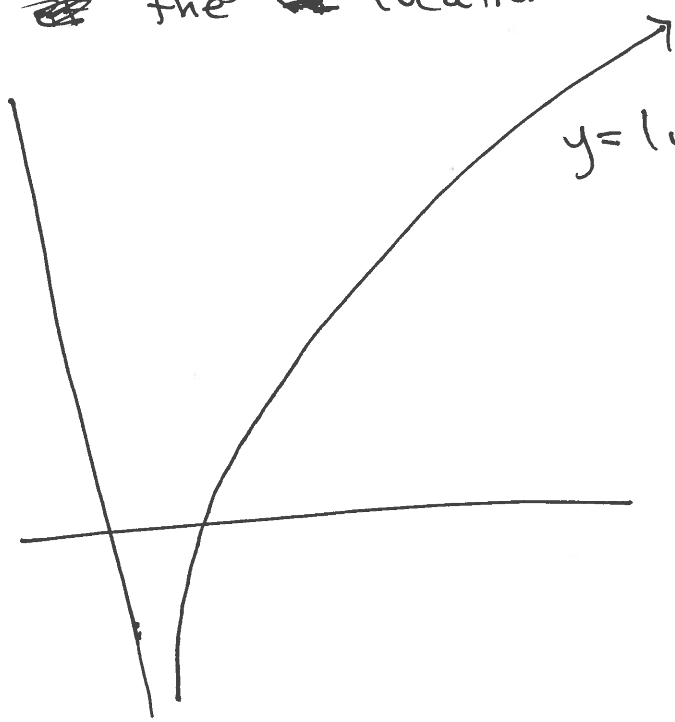
PROBLEM: Maximizing or minimizing a function that is the product of functions is yucky. (why?)
Requires the use of the product rule which can be sloppy.

SOLUTION: Take the logarithm!

DEFINITION: The log-likelihood of the random sample $X_1, \dots, X_n$, given θ, is the function

$$\ell(\theta; X_1, \dots, X_n) = \log\left(\prod_{i=1}^{n} f(\theta; X_i)\right)$$
$$= \sum_{i=1}^{n} \log\left(f(\theta; X_i)\right)$$

QN: why was this "legal"?
ANS: Monotonic increasing functions do not change the location of their extrema



STEPS OF MLE

1. Form the likelihood function L.
2. Create the log-likelihood function ℓ.
3. Compute the partial derivatives with respect to each unknown parameter.
4. Set each resulting expression equal to zero.
5. Solve the resulting equation or set of equations to identify critical points.
6. Use second-order conditions (i.e., a second derivative test, e.g., might need to use Hessian) to determine which critical point is the maximum. (Possibly need to check endpoints.)

EX: Suppose that $X_1, \ldots, X_n$ is a random sample from a normal distribution with unknown mean μ and known variance σ^2.

(1) $$L(\mu, \sigma^2; X_1, \ldots, X_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}$$

(2) $\cancel{\ell(\mu,\sigma^2)}$

$$\ell(\mu, \sigma^2; X_1, \ldots, X_n) = \sum_{i=1}^{n} \log\left\{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}\right\}$$

$$= \sum_{i=1}^{n} \left\{\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \log\left(e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}\right)\right\}$$

$$= \sum_{i=1}^{n} \left\{-\log(\sigma\sqrt{2\pi}) - \frac{(X_i-\mu)^2}{2\sigma^2}\right\}$$

$$= -n\log(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

(3) $$\frac{\partial \ell}{\partial \mu} = -\,-\frac{1}{2\sigma^2}\cdot 2\cdot\sum_{i=1}^{n}(X_i-\mu) \overset{(4)}{=} 0$$

$$\Rightarrow \left\{\sum_{i=1}^{n} X_i\right\} - n\mu = 0$$

(5)

$$\Rightarrow \mu = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

(6) $-\frac{n}{\sigma^2} < 0 \Rightarrow$ strictly concave down

$\Rightarrow \bar{X}_n = \hat{\mu}$ is a maximum

QUICK ASIDE: If you solve the same problem with σ^2 unknown, you will find the following joint solution:

$$\hat{\mu} = \bar{X}_n$$

$$\hat{\sigma}^2 = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$$

Is this the formula for sample variance?

MORAL: MLE produces possible statistics that may not be the "best" in some way, e.g., they could be biased.

We now know how to find (possibly good) statistics. We might like to know things about them — e.g., their distributions.

It turns out that the CLT (Central Limit Theorem) plays a key role in describing the distribution of statistics like $\bar{X}_n$ or $\sum_{i=1}^{n} X_i$.

<u>CENTRAL LIMIT THEOREM:</u> Suppose that $X_1, \ldots, X_n$ is a random sample, i.e., an independent and identically distributed sequence of random variables. Call the common mean μ of the random variables $\mu = \mathbb{E}[X_i]$. Call the common (finite!) variance by $\sigma^2 = Var(X_i) < \infty$. For large n (i.e., as n tends to infinity), the following results approximately hold:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

-or-

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

EX: Suppose each person in MSDS 504 (and there are $n=88$) draw an $X_i \sim Exp(\lambda)$, where $\lambda = 2$. They do so independently and we generate $\sum_{i=1}^{88} X_i$ with $X_i \sim Exp(2)$.

1. What is the approximate distribution of $\sum_{i=1}^{88} X_i$?

2. What is its mean, variance, and standard deviation?

3. Using the Empirical Rule, which symmetric interval would expect to find $\sum_{i=1}^{88} X_i$ in 95% of the time?

1. The CLT says that $\sum_{i=1}^{88} X_i$ is approximately normally distributed.

2. It has a mean equal to n $(=88)$ times $E[X_i] = \frac{1}{2}$, or $88/2 = 44$. The $Var\left(\sum_{i=1}^{88} X_i\right)$ $= \sum_{i=1}^{88} Var(X_i) = \sum_{i=1}^{88} \left(\frac{1}{2}\right)^2 = \frac{88}{4} = 22$. Thus, the standard deviation $\sigma\left(\sum_{i=1}^{88} X_i\right) = \sqrt{22}$.

3. By the Empirical Rule, approximately 95% of the time the random sum $\sum_{i=1}^{88} X_i$ will be in the interval
$$[44 - 2\sqrt{22},\ 44 + 2\sqrt{22}].$$

EX: Suppose that X_i is a randomly-drawn Stanford-Binet IQ ~~te~~ test result. Note that

$$\mathbb{E}[X_i] = 100 \quad \text{and} \quad \sigma_{X_i} = 16.$$

Suppose I take 9 MSDS students and assume they drawn from the general population. They are chosen at random and we compute the average IQ for the group.

What is the threshold such that $\bar{X}_9$ is underneath that threshold 85% of the time?

$$\Rightarrow \mathbb{E}[\bar{X}_9] = \mathbb{E}[X_i] = 100.$$

$$\Rightarrow \text{Var}(\bar{X}_9) = \frac{16^2}{9} \Rightarrow \sigma_{\bar{X}} = \sqrt{\frac{16^2}{9}} = \frac{16}{3}.$$

sometimes call this the "standard error"

$$\Rightarrow \text{qnorm}(0.85, \text{ mean} = 100, \text{ sd} = 16/3) \approx 105.52.$$

Or, using horrible table from 1968,

$$100 + 1.04\left(16/3\right) \approx 105.54.$$