In [2]:
```
display(Image(filename='./directions/0.jpg'))
```

Computer Age Statistical Inference: Exercises

Bradley Efron and Trevor Hastie
Stanford University

Many of these exercises use data used in the book. These datasets can be found on the book webpage `https://web.stanford.edu/~hastie/CASI`.

In [3]:
```
display(Image(filename='./directions/1.jpg'))
```

Chapter 1 Exercises

1. (a) Fit a cubic regression, as a function of age, to the `kidney` data of Figures 1.1 and 1.2, calculating estimates and standard errors at ages 20, 30, 40, 50, 60, 70, 80.

   (b) How do the results compare with those in Table 1.1?

2. The `lowess` curve in Figure 1.2 has a flat spot between ages 25 and 35. Discuss how one might use bootstrap replications like those in Figure 1.3 to suggest whether the flat spot is genuine or just a statistical artifact.

3. Suppose that there were no differences between AML and ALL patients for any gene, so that t in (1.6) exactly followed a student-t distribution with 70 degrees of freedom in all 7128 cases. *About* how big might you expect the largest observed t value to be? Hint: $1/7128 = 0.00014$.

4. (a) Perform 1000 nonparametric bootstrap replications of $\overline{\text{ALL}}$ (1.5). You can use program `bcanon` from the CRAN library "bootstrap" or type in the little program *Algorithm 10.1* on page 178.

   (b) Do the same for $\overline{\text{AML}}$.

   (c) Plot histograms of the results, and suggest an inference.

1

In [47]:
```
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
from statsmodels.regression.linear_model import OLS
from statsmodels.tools.tools import add_constant
```

2

One way I can think of is to Bootstrap sample only from observations from subjects with ages 25 to 35. On each bootstrap dataset I would compute a linear regression using age and and intercept. Then I would collect the value of the coefficient for age on each bootstrap iteration. Finally, I would create a histogram with the values of the age coefficients across bootstrap samples and compute the confidence interval on the age coefficients using it. If the CI ranges from negative to positive values, then I would conclude the flat spot it legitimate

3

If the distribution of the difference in the means from AML and ALL patients was indeed t with 70 degrees of freedom, then by the hint, we´d expect the largest value to be larger than the other 7127 just by chance (the probability that a random value is the largest would be 1/7128). If we look for such a value in the t distribution with 70 degrees of freedom, we get 3.826.

4

In [9]:

```
# Import libraries that we'll need...
import pandas as pd
import numpy as np
from sklearn.utils import resample
import matplotlib.pyplot as plt
from scipy import stats
import matplotlib
```

In [10]:
```
display(Image(filename='./directions/2_1.jpg'))
display(Image(filename='./directions/2_2.jpg'))
```

Chapter 2 Exercises

1. A coin with probability of heads θ is independently flipped n times, after which θ is estimated by

$$\hat{\theta} = \frac{s+1}{n+2},$$

with s equal the number of heads observed.

   (a) What are the bias and variance of $\hat{\theta}$?

   (b) How would you apply the plug-in principle to get a practical estimate of $\text{se}(\hat{\theta})$?

2. Supplement Table 2.1 with entries for trimmed means, trim proportions 0.1, 0.2, 0.3, 0.4.

3. Page 14 presents two definitions of frequentism, one in terms of probabilistic accuracy and one in terms of an infinite sequence of future trials. Give a heuristic argument relating the two.

4. Suppose that in (2.15) we plugged in $\hat{\sigma}$ to get an approximate 95% normal theory hypothesis test for $H_0 : \theta = 0$. How would it compare with the student-t hypothesis test?

5. Recompute the Neyman–Pearson alpha-beta curve in Figure 2.2, now with $n = 20$. In qualitative terms, how does it compare with the $n = 10$ curve?

1

In [12]:
```
display(Image(filename='./sol/2_1_a.jpg'))
display(Image(filename='./sol/2_1_b_1.jpg'))
display(Image(filename='./sol/2_1_b_2.jpg'))
```

1

a) · Bias calculation: First, compute $E(\hat{\theta})$

Since $\hat{\theta} = \frac{s+1}{n+2} \Rightarrow E(\hat{\theta}) = \frac{1}{n+2} \cdot E(S+1)$

$= \frac{1}{n+2} \cdot [E(S)+1]$

Expected # of successes in Binomial setting is np

$= \frac{1}{n+2} \cdot (n\theta + 1)$

$= \frac{n\theta+1}{n+2}$

Then, compute $\theta - E(\hat{\theta})$

$$\text{Bias} = \theta - \left[\frac{n\theta}{n+2} + \frac{1}{n+2}\right] = \theta\Big(\underbrace{\frac{n+2}{n+2}}_{=1} - \frac{n}{n+2}\Big) - \frac{1}{n+2}$$

$$= \frac{2\theta - 1}{n+2}$$

· Variance Calculation

$$V(\hat{\theta}) = V\left(\frac{s+1}{n+2}\right) = \frac{1}{(n+2)^2} \cdot V(S+1)$$

$$= \frac{V(S)}{(n+2)^2} = \frac{n \cdot \theta \cdot (1-\theta)}{(n+2)^2}$$

b) $se(\hat{\theta})$ is just $\sqrt{V(\hat{\theta})}$

Thus, $se(\hat{\theta}) = \frac{[n \cdot \theta \cdot (1-\theta)]^{1/2}}{n+2}$

- Since we don't know θ, we use our estimator $\hat{\theta} = \frac{\quad}{n+2}$ in its place (this is the plug-in principle) and get

$$se(\hat{\theta}) = \frac{\left[n\left(\frac{s+1}{n+2}\right)\cdot\left(1-\frac{s+1}{n+2}\right)\right]^{1/2}}{n+2}$$

and then, of course, you can further simplify it.

2

In [2]:
```python
# Import libraries that we'll need...
import pandas as pd
import numpy as np
from sklearn.utils import resample
from scipy.stats.mstats import winsorize
import matplotlib.pyplot as plt
from scipy import stats
import matplotlib
```

3

In [13]:
```python
display(Image(filename='./sol/2_3.jpg'))
```

(3) Frequentism assigns the properties of the estimator (like its bias & variance) to the value of it we got from our very specific sample. The reason for doing this (and the link to the repeated sampling assumption), is that your very specific sample is a random one from the infinite possible samples.

4

```
In [14]: display(Image(filename='./sol/2_4.jpg'))
```

(4) The test would asymptotically converge to the t-student's test for larger n_1+n_2, as sample increase would be accompanied by asymptotic convergence of $\hat{\sigma}$ to σ.

5

```
In [1]: import numpy as np
        from scipy.stats import norm
        import matplotlib.pyplot as plt
```

In [6]:
```
display(Image(filename='./directions/3.jpg'))
```

Chapter 3 Exercises

1. Suppose the parameter μ in the Poisson density (3.3) is known to have prior density $e^{-\mu}$. What is the posterior density of μ given x?

2. In Figure 3.1, suppose the doctor had said "1/2, 1/2" instead of "1/3, 2/3". What would be the answer to the physicist's question?

3. Let X be binomial,

$$\Pr_\pi\{X = x\} = \binom{n}{x}\pi^x(1-\pi)^{n-x} \qquad \text{for } x = 0, 1, \ldots, n.$$

   What is the Fisher information $\mathcal{I}_\pi$ (3.16)? How does $\mathcal{I}_\pi$ relate to the estimate $\hat{\pi} = x/n$?

4. (a) Run the following simulation 200 times:
   - $x_i \overset{\text{ind}}{\sim} \mathcal{N}(\mu_i, 1)$ for $i = 1, 2, \ldots, 500$
   - $\mu_i = 3i/500$
   - i_{max} = index of largest x_i
   - $d = x_{i_{\text{max}}} - \mu_{i_{\text{max}}}$

   (b) Plot the histogram of the 200 d values.

   (c) What is the relation to Figure 3.4?

5. Give a brief nontechnical explanation of why $x_{610} = 5.29$ was likely to be an overestimate of θ_{610} in Figure 3.4.

6. Given prior density $g(\mu)$ and observation $x \sim \text{Poi}(\mu)$, you compute $g(\mu \mid x)$, the posterior density of μ given x. Later you are told that x could only be observed if it were greater than 0. (Table 6.2 presents an example of this situation.) Does this change the posterior density of μ given x?

1

In [15]:
```
display(Image(filename='./sol/3_1.jpg'))
```

(1) $f_\mu(x) = \dfrac{e^{-\mu} \cdot \mu^x}{x!}$

$\pi(\mu) = e^{-\mu}$

Will assume x is one observation, as vector notation not used explicitly in the problem.

$$P(\mu|x) \propto \pi(\mu) \cdot f_\mu(x)$$
$$\propto e^{-\mu} \cdot e^{-\mu} \cdot \mu^x$$
$$\propto e^{-2\mu} \cdot \mu^x$$

Compare this with the Gamma distribution

$y \sim Ga(\alpha, \beta)$

then $f(y) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \cdot e^{-\beta y}$

$$\propto y^{\alpha-1} \cdot e^{-\beta y}$$

Dropping the constants & just keeping the kernel

Thus, $y = \mu$, $\beta = 2$ and $\alpha - 1 = x$
$\Rightarrow \alpha = x+1$

and $P(\mu|x) \propto e^{-2\mu} \cdot \mu^x \propto Ga(x+1, 2)$

2

In [16]:

```
display(Image(filename='./sol/3_2_1.jpg'))
display(Image(filename='./sol/3_2_2.jpg'))
```

2

Sonogram

Twins are:

	SS	Diff	
I	1/2	0	1/2
F	1/4	1/4	1/2

Then

$$\frac{g(I \mid s)}{g(F \mid s)} = \cancel{\frac{1/2}{1/2}} \cdot \frac{1}{1/2} = 2$$

Identical twins would be twice as likely as fraternals after sonogram results.

3

In [17]:
```
display(Image(filename='./sol/3_3_1.jpg'))
display(Image(filename='./sol/3_3_2.jpg'))
```

3 Going step-by-step, let's first compute the score function.

② is below.

$$\dot{\ell}_{\underline{x}}(\pi) = \frac{\partial}{\partial \pi}\left\{\log\left[\binom{n}{x}\right] + x \cdot \log(\pi) + (n-x)\cdot \log(1-\pi)\right\}$$

$$= \cancel{\frac{\partial}{\partial \pi} \log\binom{n}{x}}^{\,0} + \frac{x}{\pi} - \frac{(n-x)}{1-\pi}$$

$$= \frac{x\cdot(1-\pi) - (n-x)\cdot \pi}{\pi(1-\pi)} = \frac{x - \cancel{x\pi} - n\pi + \cancel{x\pi}}{\pi(1-\pi)}$$

Now, using the definition of Fisher Information:

$$\mathcal{I}_\pi = \sum_{x=0}^{n}\binom{n}{x}\cdot \pi^x \cdot (1-\pi)^{n-x}\cdot\left(\frac{x-n\pi}{\pi(1-\pi)}\right)^2$$

$$= \frac{1}{(\pi(1-\pi))^2}\cdot \sum_{x=0}^{n}\binom{n}{x}\cdot\pi^x\cdot(1-\pi)^{n-x}\cdot(x-n\pi)^2$$

$$= \frac{1}{(\pi(1-\pi))^2}\cdot \underbrace{\sum_{x=0}^{n}\binom{n}{x}\cdot\pi^x\cdot(1-\pi)^{n-x}\cdot(x-E(x))^2}_{\text{This is the definition of the variance for a Binomial}(n,\pi)\text{ distr, which is known to be equal to } n\cdot\pi\cdot(1-\pi)}$$

$$= \frac{n\cancel{\pi}\,\cancel{(1-\pi)}}{(\pi\cdot(1-\pi))^{\cancel{2}}} = \frac{n}{\pi(1-\pi)}.$$

4

```python
In [13]: import pandas as pd
         import numpy as np
         from scipy.stats import norm
```

5

```python
In [18]: display(Image(filename='./sol/3_5.jpg'))
```

(5) The Effect of gene 610 measured on our sample, X_{610} would have been a different number in other samples from the same population. If we think about θ_{610}, the true effect size vs. $X^{(i)}_{610}$ for different samples, it would have looked like:

The fact that we chose X_{610} because it was large vs. the other gene effects, makes it more likely to be one of the larger measurements across samples, biasing the estimation of its effect.

6

In [19]:
```
display(Image(filename='./sol/3_6_1.jpg'))
display(Image(filename='./sol/3_6_2.jpg'))
```

6 Posterior of μ given x doesn't change, as our belief of what the parameter μ is only depends on the data (x) through the likelihood

Now, if we let y: "Information telling us only $x \geq 1$ can be observed", then posterior μ given x & y would change, as the new information on y would modify the likelihood func

In [7]:
```
display(Image(filename='./directions/4.jpg'))
```

Chapter 4 Exercises

1. (a) Verify formula (4.10).
   (b) Why isn't the formula for $\hat{\sigma}$ the one generally used in practice?
2. Draw a schematic graph of $\dot{l}_x(\theta)$ versus θ. Use it to justify (4.25).
3. You observe $x_1 \sim \mathrm{Bin}(20, \theta)$ and, independently, $x_2 \sim \mathrm{Poi}(10 \cdot \theta)$. Numerically compute the Cramér–Rao lower bound (4.33). Hint: Fisher information adds for independent observations.
4. A coin with unknown probability of heads θ is flipped n_1 times, yielding x_1 heads; then it is flipped another x_1 times, yielding x_2 heads.
   (a) What is an intuitively plausible estimate of θ?
   (b) What Fisherian principle have you invoked?
5. Recreate a version of Figure 4.3 based on 1000 permutations.
6. A one-parameter family of densities $f_\theta(x)$ gives an observed value x. Statistician A computes the MLE $\hat{\theta}$. Statistician B uses a flat prior density $g(\theta) = 1$ to compute $\bar{\theta}$, the Bayes posterior expectation of θ given x. Describe the relationship between the two methods.

1

In [20]:

```
display(Image(filename='./sol/4_1_1.jpg'))
display(Image(filename='./sol/4_1_2.jpg'))
display(Image(filename='./sol/4_1_3.jpg'))
```

Chapter 4 Sols + 5.1 Sols

1) a) $\rightarrow$

$$f_x(\theta,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}}$$ (Density for 1 observation)

$$f_{\underline{x}}(\theta,\sigma) = (2\pi\sigma^2)^{-n/2} \cdot \exp\left\{\sum_{i=1}^{n} -\frac{1}{2}\frac{(x_i-\theta)^2}{\sigma^2}\right\}$$ (likelihood of sample)

$$l_{\underline{x}}(\theta,\sigma) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\cdot\sum_{i=1}^{n}\frac{(x_i-\theta)^2}{\sigma^2}$$ (log-likelihood of sample)

Optimality condition

$$\frac{\partial l_x}{\partial \theta} = \sum_{i=1}^{n}\frac{(x_i-\theta)}{\sigma^2} = 0 \Rightarrow \sum_{i=1}^{n} x_i - n\theta = 0$$

Derivative w.r.t. θ

$$\Rightarrow \boxed{\hat{\theta}^{mle} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{X}}$$ (mean estimator)

Optimality Condition

$$\frac{\partial l_x}{\partial \sigma} = -n\cdot\frac{\cancel{2\pi}\,\cancel{\sigma}}{\cancel{2\pi}\,\sigma^{\cancel{2}}} + \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{\sigma^3} = 0$$

Derivative w.r.t. σ

$$\Rightarrow \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{\sigma^3} = \frac{n}{\sigma}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\theta)^2}{n}$$ (Estimator for variance.)

$$\boxed{\sigma = \left[\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n}\right]^{1/2}}$$ (Using plug-in principle $\hat{\theta}=\overline{X}$ and taking square root.)

b) $\rightarrow$

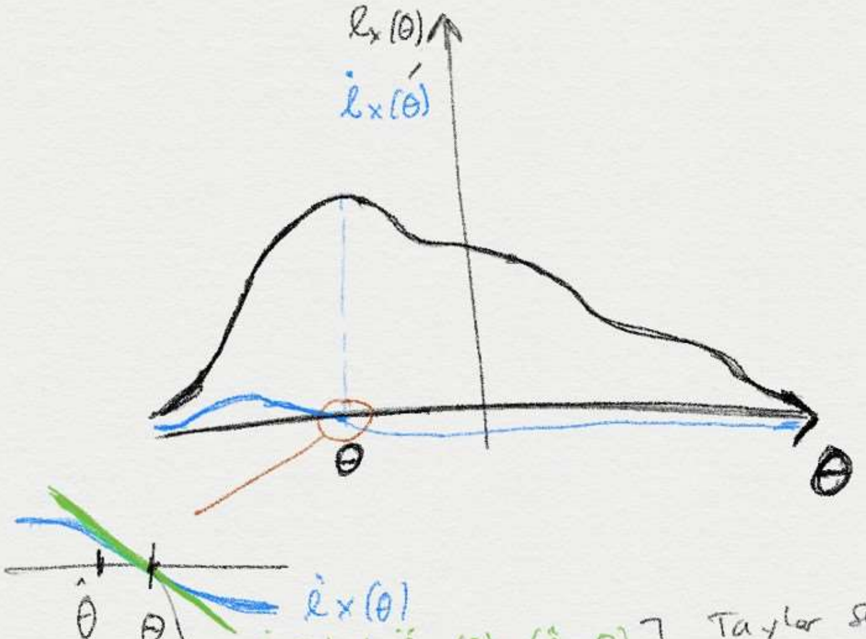
Estimator for variance only takes into account current sample when maximizing likelihood. It ignores the fact that $\overline{X}$, our plug-in estimate for θ, results in loss of 1 degree of freedom for the residuals.

2

In [21]:
```
display(Image(filename='./sol/4_2_1.jpg'))
display(Image(filename='./sol/4_2_2.jpg'))
display(Image(filename='./sol/4_2_3.jpg'))
```

2 Given a log-likelihood function $\ell_X(\theta)$, we observe that score function $\dot{\ell}_X(\theta)$ is zero at the value of θ where $\ell_X(\theta)$ is maximized.

zero at the value of θ where $\ell_X(\theta)$ is maximized.



$\dot{\ell}(\theta) + \ddot{\ell}_x(\theta)\cdot(\hat{\theta}-\theta)$ } Taylor series around the parameter θ

$$\Rightarrow 0 = \dot{\ell}(\theta) + \ddot{\ell}_x(\theta)\cdot(\hat{\theta}-\theta)$$

$$(\theta-\hat{\theta})\,\ddot{\ell}_x(\theta) = \dot{\ell}_x(\theta)$$

$$\hat{\theta} = \theta - \frac{\dot{\ell}_x(\theta)}{\ddot{\ell}_x(\theta)} = \theta - \frac{\dot{\ell}_x(\theta)/n}{\ddot{\ell}_x(\theta)/n}$$

3

In [23]:

```
display(Image(filename='./sol/4_3_1.jpg'))
display(Image(filename='./sol/4_3_2.jpg'))
display(Image(filename='./sol/4_3_3.jpg'))
display(Image(filename='./sol/4_3_4.jpg'))
display(Image(filename='./sol/4_3_5.jpg'))
display(Image(filename='./sol/4_3_6.jpg'))
```

③ Cramér-Rao Bound

$$V_\theta\{\tilde{\theta}\} \geq \frac{1}{I_\theta}$$

Total Fisher Information, if observations are iid then $= n \cdot I_\theta^{(1)}$ where $I_\theta^{(1)}$ is the Fisher Inf. of 1 observation.

$X_1 \sim Bin(20, \theta)$ sampled Indep.

$X_2 \sim Poi(10 \cdot \theta)$

$$f(x_1) = \binom{n}{x_1} \cdot \theta^{x_1} \cdot (1-\theta)^{n-x_1}$$

$$l_{x_1}(\theta) = \log\left[\binom{n}{x_1}\right] + x_1 \log(\theta) + (n-x_1)\log(1-\theta)$$

$$\dot{l}_{x_1}(\theta) = \frac{x_1}{\theta} + \frac{(n-x_1)}{1-\theta}$$

$$\ddot{l}_{x_1}(\theta) = -\frac{(n-x_1)}{(1-\theta)^2} - \frac{x_1}{\theta^2}$$

$$I_\theta = -E\left[\ddot{l}_{x_1}(\theta)\right] = \frac{(n - n \cdot \theta)}{(1-\theta)^2} + \frac{n\cancel{\theta}}{\theta^{\cancel{2}}}$$

$$= \frac{n\cancel{(1-\theta)}}{(1-\theta)^{\cancel{2}}} + \frac{n}{\theta} = \frac{\cancel{n\theta} + n - \cancel{n\theta}}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)}$$

$$= \boxed{\frac{20}{\theta(1-\theta)}}$$

Let $\lambda = 10 \cdot \theta$

$$f_{x_2}(\lambda) = \frac{e^{-\lambda} \cdot \lambda^{x_2}}{x_2!}$$

$$\ell_{x_2}(\lambda) = -\lambda + x_2 \log(\lambda) - \log(x_2!)$$

$$\dot{\ell}_{x_2}(\lambda) = -1 + \frac{x_2}{\lambda}$$

$$\ddot{\ell}_{x_2}(\lambda) = \frac{-x_2}{\lambda^2}$$

Then, taking the expectation & substituting θ back...

$$I_\theta = -E\left[\ddot{\ell}_{x_2}(\theta)\right] = -E\left[\frac{-x_2}{(10 \cdot \theta)^2}\right] = \cancel{-}\frac{\cancel{-(10\cdot\theta)}}{(10\cdot\theta)^{\cancel{2}}} = \frac{1}{10 \cdot \theta}$$

Finally, $I_{\theta|x_1,x_2} = \frac{20}{\theta(1-\theta)} + \frac{1}{10\theta}$ — Total Fisher Inf. of both observations.

Finally, $I_{\theta|x_1,x_2} = \frac{20}{\theta(1-\theta)} + \frac{1}{10\theta}$ — Total Fisher Inf. of both observations.

Let's now compute numerically, as requested by problem. Since θ is the success probability of a Binomial distribution, we know $0 \le \theta \le 1$. Making a grid in increments of 0.1, we get

θ	$I_{\theta\|x_1,x_2}$	$1/I_\theta$	Information gain

θ	$I_0 \mid x_1, x_2$	$1/I_0$
0.1	223	0.0044
0.2	125	0.0080
0.3	95	0.0104
0.4	84	0.0119
0.5	96	0.0125
0.6	84	0.0119
0.7	95	0.105
0.8	125	0.0080
0.9	222	0.0045

Information gain is greater at the edges !!

Variance of any unbiased estimator $\geq 1/I_0$ for this sample.

4

In [27]:
```python
display(Image(filename='./sol/4_4.jpg'))
```

4

a) $\frac{x_1 + x_2}{n_1 + x_1}$, the total amount of successes over the total amount of trials.

b) Conditional Inference, because we considered the outcomes of the individual samples, regardless of how the second sample was generated (i.e, ignoring the distribution of x_1).

5

In [7]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import cauchy
from scipy.optimize import minimize
from statsmodels.regression.linear_model import OLS
```

6

In [28]:
```
display(Image(filename='./sol/4_6_1.jpg'))
display(Image(filename='./sol/4_6_2.jpg'))
```

6 Since the posterior computed by statistician B is proportional to the likelihood,

$$P(\theta \mid x) \propto f_x(\theta) \cdot g(\theta)$$

$$\propto f_x(\theta)$$

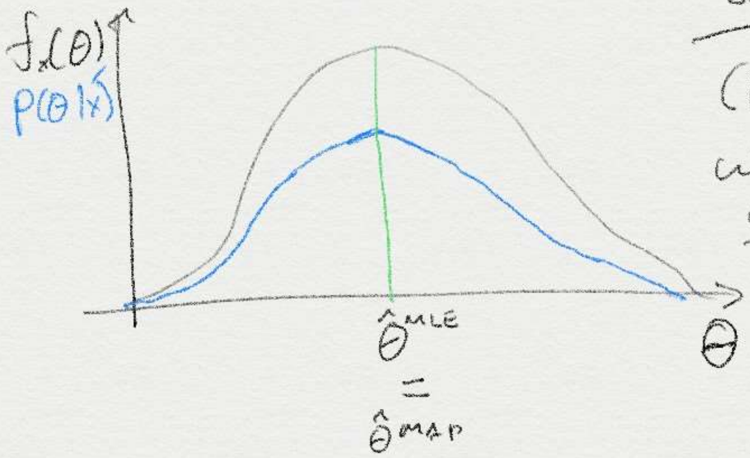
which is the function statistician A is trying to maximize. Thus, under a flat prior, they have the

which is the function statistician A is trying to maximize. Thus, under a flat prior, they have the

same maximizer.

$(\hat{\theta}^{MLE} = \hat{\theta}^{MAP})$

where $\hat{\theta}$ is the maximum a posteriori estimate.



In [11]:

```
display(Image(filename='./directions/5_1.jpg'))
display(Image(filename='./directions/5_2.jpg'))
```

Chapter 5 Exercises

1. Suppose $X \sim \text{Poi}(\mu)$ where μ has a $\text{Gam}(\nu, 1)$ prior (as in Table 5.1).

   (a) What is the marginal density of X?

   (b) What is the conditional density of μ given $X = x$?

2. X is said to have an "F distribution with degrees of freedom ν_1 and ν_2", denoted $F_{\nu_1,\nu_2}(x)$, if

$$X \sim \frac{\nu_2}{\nu_1} \frac{\text{Gam}(\nu_1, \sigma)}{\text{Gam}(\nu_2, \sigma)},$$

   the two gamma variates being independent. How does the F distribution relate to the beta distribution?

3. Draw a sample of 1000 bivariate normal vectors $x = (x_1, x_2)'$, with

$$x \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right).$$

   (a) Regress x_2 on x_1, and numerically check (5.18).

   (b) Do the same regressing x_1 on x_2.

4. Suppose $x \sim \mathcal{N}_p(\mu, \mathbf{\Sigma})$ as in (5.14), with $\mathbf{\Sigma}$ a *known* $p \times p$ matrix. Use (5.26) to directly calculate the information matrix $\mathcal{I}_\mu$. How does this relate to (5.27)?

5. Draw the equivalent of Figure 5.5 for $x \sim \text{Mult}_3(5, \boldsymbol{\pi})$.

6. If $x \sim \text{Mult}_L(n, \boldsymbol{\pi})$, use the Poisson trick (5.44) to approximate the mean and variance of x_1/x_2. (Here we are assuming that $n\pi_2$ is large enough to ignore the possibility $x_2 = 0$.) Hint: In notation (5.41),

$$\frac{S_1}{S_2} \doteq \frac{\mu_1}{\mu_2}\left(1 + \frac{S_1 - \mu_1}{\mu_1} - \frac{S_2 - \mu_2}{\mu_2}\right).$$

7. Show explicitly how the binomial density $\text{bi}(12, 0.3)$ is an exponential tilt of $\text{bi}(12, 0.6)$.

In [29]:
```
display(Image(filename='./sol/5_1_1.jpg'))
display(Image(filename='./sol/5_1_2.jpg'))
display(Image(filename='./sol/5_1_3.jpg'))
```

Chapter 5, 1.

$X \sim Poi(\mu)$, $\pi(\mu) \sim Ga(V, 1)$

a) Before getting the marginal density of X, let's get the joint density for x, μ.

$$f(x, \mu | V) = f_x(\mu) \cdot \pi(\mu) \quad \text{(using conditional Probability)}$$

$$= \frac{\mu^x \cdot e^{-\mu}}{x!} \cdot \frac{\mu^{V-1} \cdot e^{-\mu/1}}{1^V \, \Gamma(V)}$$

Recall $\Gamma(y+1) = y!$

$$\frac{\mu^{(V+x)-1} \cdot e^{-2\mu}}{\Gamma(x-1) \cdot \Gamma(V)}$$

Now, get marginal density $f(x)$ by integrating over μ.

$$f(x|V) = \frac{1}{\Gamma(x-1)} \cdot \int_0^\infty \frac{\mu^{(V+x)-1} \cdot e^{-\mu/(1/2)}}{\Gamma(V)} d\mu$$

$$f(x|V) = \frac{(1/2)^{V+x}}{\Gamma(x-1)} \cdot \underbrace{\int_0^\infty \frac{\mu^{(V+x)-1} \cdot e^{-\mu(1/2)}}{(1/2)^{V+x} \cdot \Gamma(V)} d\mu}_{\text{Gamma density } Ga(V+x, 1/2)}$$

$$\boxed{\Rightarrow f(x|V) = \frac{(1/2)^{V+x}}{\Gamma(x-1)}}$$

b) Using Bayes Rule

$$f(\mu | x, v) = \frac{f(\mu, x | v)}{f(x|v)}$$

So just substitute them & simplify ☺.