STAT. INFERENCE 453/553 02/21/03

~~TEST 1 ONE WEEK FROM TODAY 02/28/03~~

~~REVIEW SESSION ON WED 26~~

TODAY: MORE LIKELIHOOD AND POINT ESTIMATION

EX: NORMAL CASE. LET $X_1, X_2, \ldots X_n$ BE iid RVs SUCH THAT $X_i \sim N(\mu, 1)$. WHAT IS THE LIKELIHOOD FUNCTION FOR μ?

$$L(\mu|\underline{x}) = \prod_{i=1}^{n} f(x_i|\mu)$$

IN THIS CASE, $\sigma = 1$

$$\prod_{i=1}^{n} f(x_i|\mu) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right) \exp\left(-\frac{1}{2/n}(\mu-\bar{x})^2\right).$$

$$\left(\prod_{i=1}^{n} f(x_i|\mu) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right) \text{ + REMEMBER}\right).$$

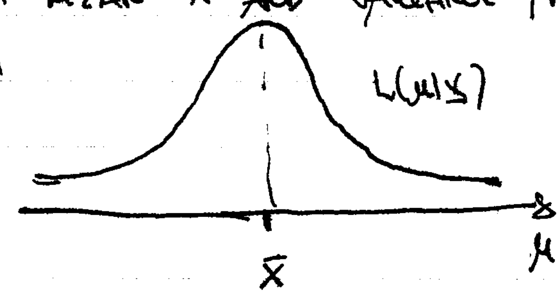
AS A FUNCTION OF μ,

$\left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)$ IS CONSTANT $(C(\underline{x}))$

$$\Rightarrow L(\mu|\underline{x}) = C(\underline{x}) \underbrace{\exp\left(-\frac{1}{2/n}(\mu-\bar{x})^2\right)}_{1}$$

NOTICE THAT 1 LOOK LIKE THE KERNEL OF A NORMAL DENSITY (ON μ!) WITH MEAN $\bar{x}$ AND VARIANCE $1/n$.

GRAPH



THIS LIKELIHOOD DOES NOT INTEGRATE TO ONE. BECAUSE THE CONSTANT IS $C(\underline{x})$ AND NOT $\frac{1}{\sqrt{2\pi(1/n)}}$

CERTAINLY, IF WE MAKE

$$L^*(\mu|\underline{x}) = c(\underline{x})^{-1} \frac{1}{\sqrt{2\pi\, 1/n}} L(\mu|\underline{x}) = \frac{1}{\sqrt{2\pi\, 1/n}} \exp\left(\frac{-1}{2/n}(\mu-\bar{x})^2\right)$$

THIS A pdf ON $\mathcal{M}$ (THE FIDUCIAL DIST ON μ).

WE COULD COMPUTE THINGS LIKE:

$$\Pr[a \leq \mu \leq b] = \Pr\left(\frac{a-\bar{x}}{\sqrt{1/n}} \leq Z \leq \frac{b-\bar{x}}{\sqrt{1/n}}\right); \; Z \sim N(0,1)$$

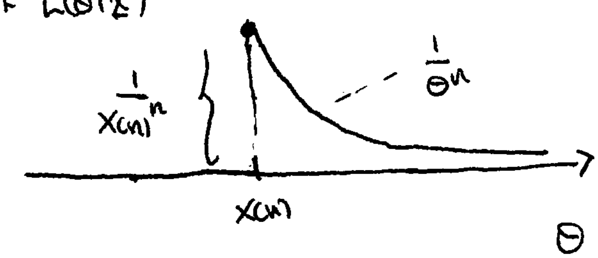
Ex:

LET $X_1, X_2, \ldots X_n$ BE iid RVs SUCH THAT EACH $X_i \sim U(0,\theta)$

$$\Rightarrow f(x_i|\theta) = \frac{1}{\theta} \underset{(0,\theta)}{I}(x_i), \quad \Rightarrow L(\theta|\underline{x}) = \prod_{i=1}^{n} \frac{1}{\theta} \underset{(0,\theta)}{I}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^{n} \underset{(0,\theta)}{I}(x_i)$$

$$= \frac{1}{\theta^n} \underset{(0,\infty)}{I}(x_{(1)}) \underset{(x_{(n)},\infty)}{I}(\theta) = c \cdot \frac{1}{\theta^n} \underset{(x_{(n)},\infty)}{I}(\theta)$$

GRAPH OF $L(\theta|\underline{x})$



TO TRANSFORM TO A DENSITY.

$$\int_{x_{(n)}}^{\infty} \frac{1}{\theta^n} d\theta = \frac{-\theta^{-(n+1)}}{(n+1)} \Bigg|_{x_{(n)}}^{\infty} = \frac{(x_{(n)})^{-(n+1)}}{(n+1)}$$

$$\Rightarrow L^*(\theta|\underline{x}) = (n+1)\,(x_{(n)})^{-(n+1)} \frac{1}{\theta^n} \underset{(x_{(n)},\infty)}{I}(\theta)$$ IS THE FIDUCIAL pdf ON θ.

DO YOU RECOGNIZE THIS pdf? IT IS A PARETO pdf WITH $\alpha = x_{(n)}$ AND $\beta = n+1$

POINT ESTIMATION

IDEA: PROPOSE A STATISTIC $W(X_1, X_2 .\ X_n)$ THAT IS A GOOD ESTIMATOR OF THE POINT θ. (NOTATION: $\hat{\theta}(\underline{X})$; $\hat{\theta}(X_1 .\ X_n)$, $\tilde{\theta}$)

HOW TO PROPOSE ? WHAT IS GOOD ?

METHODS OF POINT ESTIMATION

MAXIMUM LIKELIHOOD ESTIMATION.

FOR EACH SAMPLE POINT $\underline{x}$, LET $\hat{\theta}(\underline{x})$ BE THE VALUE ~~that~~ AT ~~[illegible]~~ WHICH $L(\theta|\underline{x})$ ATTAINS ITS MAXIMUM. (GLOBAL)

THE STATISTIC $T = \hat{\theta}(\underline{X})$ IS A MAXIMUM LIKELIHOOD ESTIMATOR FOR θ.

EX: FOR X'S $\sim U(0,\theta)$, THE MAXIMUM LIKELIHOOD ESTIMATOR IS $\hat{\theta}(\underline{X}) = X_{(n)} = \max\{X_1, X_2 .\ X_n\}$.

EX: FOR X'S $\sim N(\mu, 1)$ THE LIKELIHOOD ~~IS MAXIMAL~~ REACHES ITS MAXIMUM AT ~~[illegible]~~ $\hat{\mu} = \bar{x}$

ANOTHER WAY TO SHOW THIS:

$$L(\mu|\underline{x}) = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$\log L(\mu|\underline{x}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{d \log L(\mu|\underline{x})}{d\mu} = 0 + 2\left(\frac{1}{2}\right)\sum_{i=1}^{n}(x_i-\mu)$$

MAKE DERIVATIVE EQUAL TO ZERO

$$\frac{d \log L(\mu|\underline{x})}{d\mu} = 0 \Rightarrow \sum_{i=1}^{n}(x_i-\mu) = 0 \Rightarrow \sum_{i=1}^{n} x_i = n\mu$$

$\Rightarrow$ ~~[illegible]~~ $\mu = \bar{x}$ IT IS A ZERO OF THE FIRST DERIVATIVE.

$$\frac{d^2 \log L(\mu|\underline{X})}{d\mu^2} = -n < 0 \Rightarrow \hat{\mu} = \bar{X} \text{ IS THE MLE FOR } \mu.$$

IN GENERAL, IF WE HAVE A LIKELIHOOD FUNCTION $L(\theta|\underline{X})$ DIFFERENTIABLE ON θ, POSSIBLE CANDIDATES FOR THE MLE ARE VALUES $(\theta_1, \theta_2, \ldots \theta_k)$ SUCH THAT

$$\frac{\partial L(\theta|\underline{X})}{\partial \theta_i} = 0, \quad i = 1, 2 \ldots k$$

(you CAN USE $\log L(\theta|\underline{X})$ INSTEAD OF $L(\theta|\underline{X})$, SINCE $\log(\cdot)$ IS A ONE TO ONE FUNCTION)

NOTICE THAT IN THE $U(0,\theta)$ EX WE CANNOT APPLY THIS, THE LIKELIHOOD FUNCTION IS NOT DIFFERENTIABLE FOR ALL VALUES OF θ.

EX. SUPPOSE WE HAVE AN URN WITH BLACK AND RED BALLS AND WE EXTRACT 25 BALLS WITH REPLACEMENT FROM THIS URN. AND 9 OF THESE BALLS ARE BLACK. LET p = PROB. BALL IS BLACK IN ANY EXTRACTION. WHAT IS OUR MLE ESTIMATE FOR p? INTUITION TELLS: $\hat{p} = 9/25$

FORMALLY $X \sim \text{BINOMIAL}(25, p)$ OBSERVED VALUE OF X IS 9.

$$L(p; x=9) = \binom{25}{9} p^9 (1-p)^{16} \Rightarrow \log L(p, x=9) = \log \binom{25}{9} + 9 \log p + 16 \log(1-p)$$

$$\frac{d \log L(p, x=9)}{dp} = 9\left(\frac{1}{p}\right) - 16\left(\frac{1}{1-p}\right) = 0 \Rightarrow 9(1-p) = 16p$$

$$\Rightarrow p = 16/25$$

(CHECK THAT $\frac{d^2 \log L(p, x=9)}{d^2 p} < 0$)

OBS: THE MAXIMUM LIKELIHOOD ESTIMATOR IS NOT UNIQUE

IF $X_1, X_2, \ldots X_n$, $X_i \sim \text{Bern}(\theta)$ RVs THEN

$$L(\theta|\underline{X}) = \theta^{\sum X_i}(1-\theta)^{n-\sum X_i} \qquad 0 \leq \theta \leq 1$$

$$l = \log L(\theta|\underline{X}) = \sum_{i=1}^n x_i \log(\theta) + \left(n - \sum_{i=1}^n x_i\right) \log(1-\theta)$$

$$\Rightarrow \frac{dl}{d\theta} = \frac{\sum x_i}{\theta} - \frac{(n - \sum x_i)}{(1-\theta)} = 0 \Leftrightarrow (1-\theta)\sum_{i=1}^n x_i = \theta\left(n - \sum_{i=1}^n x_i\right)$$

$$\Leftrightarrow \sum_{i=1}^n x_i = \theta n \Rightarrow \theta = \bar{X}$$

$$\frac{d^2 l}{d\theta^2} = -\frac{\sum x_i}{\theta^2} - \frac{(n - \sum x_i)}{(1-\theta)^2} < 0 \qquad \hat{\theta} = \bar{X} \text{ IS THE MLE}$$

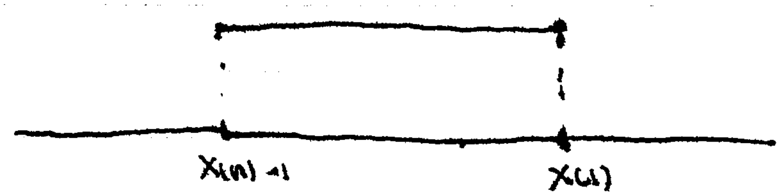
THE MLE IS NOT UNIQUE

EXAMPLE: SUPPOSE $X_1, X_2, \ldots X_n$ ARE iid $U(\theta, \theta+1)$, $i = 1, 2, \ldots n$

$$f(x_i|\theta) = I_{(\theta,\theta+1)}(x_i) \Rightarrow$$

$$L(\theta|\underline{X}) = \prod_{i=1}^n I_{(\theta,\theta+1)}(x_i) = I_{(\theta,\infty)}(x_{(1)})\, I_{(-\infty,\theta+1)}(x_{(n)}) = I_{(x_{(n)}-1,\, x_{(1)})}(\theta)$$

PLOT OF $L(\theta|\underline{X})$



THE MLE FOR θ IS ANY VALUE BETWEEN $(X_{(n)}-1, X_{(1)})$.

INVARIANCE PROPERTY

USUALLY OUR DISTRIBUTION IS INDEXED BY θ $(f(x|\theta))$

WE MIGHT BE INTERESTED IN FINDING AN ESTIMATOR FOR $\tau(\theta)$, SOME FUNCTION OF θ.

INV. PROPERTY: IF $\hat{\theta}$ IS THE MLE OF THEN $\tau(\hat{\theta})$ IS THE MLE FOR $\tau(\theta)$.

BACK TO BLLI(Θ) EXAMPLE. SUPPOSE WE WANT TO ESTIMATE $\tau(\theta) = \log\left(\theta/(1-\theta)\right)$ RATHER THAN JUST Θ. SINCE $\hat{\Theta} = \bar{X}$ IS THE MLE FOR Θ $\Rightarrow$ $\tau(\hat{\theta}) = \log\left(\bar{x}/(1-\bar{x})\right)$ IS THE MLE FOR $\tau(\theta)$.

HOW DO WE FORMALIZE THIS ?

IF $\tau(\theta)$ IS A ONE-TO-ONE FUNCTION, THERE IS NO PROBLEM.

LET $\eta = \tau(\theta) \Leftrightarrow \theta = \tau^{-1}(\eta)$. THE LIKELIHOOD FOR η IS

$$L^*(\eta|\underline{X}) = \prod_{i=1}^{n} f(x_i|\tau^{-1}(\eta)) = L(\tau^{-1}(\eta)|\underline{x}) = L(\theta|\underline{x})$$

$\Rightarrow$

$$\sup_{\eta} L^*(\eta|\underline{x}) = \sup_{\eta} L(\tau^{-1}(\eta)|\underline{x}) = \sup_{\theta} L(\theta|\underline{x})$$

THUS, THE MAXIMUM OF η IS ATTAINED AT $\eta = \tau(\hat{\theta})$.

THE PROBLEM IS WHEN WE DEAL WITH FUNCTIONS THAT ARE NOT ONE-TO-ONE (EX μ^2 IN THE NORMAL MODEL)

IF $\hat{\Theta}$ IS THE MLE FOR Θ $\Rightarrow$ SINCE $\tau(\cdot)$ IS NOT ONE-TO-ONE, WE MAY FIND ANOTHER VALUE Θ_0 SUCH THAT $\tau(\hat{\theta}) = \tau(\theta_0)$

WE DEFINE THE "INDUCED" LIKELIHOOD FUNCTION FOR $\eta = \tau(\theta)$

$$L^*(\eta|\underline{x}) = \sup_{\{\theta:\tau(\theta)=\eta\}} L(\theta|\underline{x})$$

THE VALUE $\hat{\eta}$ THAT MAXIMIZES $L^*(\eta|\underline{X})$ IS CALLED THE MAXIMUM LIKELIHOOD ESTIMATOR FOR η

BY TAKING "SUP", WE ~~GET~~ RID OF THE PROBLEM OF LACK OF UNIQUENESS IN THE TRANSFORMATION.

WE MUST SHOW THAT $L^*(\hat{\eta}|\underline{x}) = L^*(\tau(\hat{\theta})|\underline{x})$

$\Rightarrow \tau(\hat{\theta})$ IS A MLE FOR η.

$$L^*(\hat{\eta}|\underline{x}) = \sup_{\eta} \sup_{\{\theta:\tau(\theta)=\eta\}} L(\theta|\underline{x}) = \sup_{\theta} L(\theta|\underline{x}) = L(\hat{\theta}|\underline{x})$$

FURTHER, by DEF.

$$L^*[\tau(\hat{\theta})|\underline{x}] = \sup_{\{\theta:\tau(\theta)=\tau(\hat{\theta})\}} L(\theta|\underline{x}) = L(\hat{\theta}|\underline{x})$$ BECAUSE

$\hat{\theta}$ BELONGS TO THE SET $\{\theta : \tau(\theta) = \tau(\hat{\theta})\}$

CONCLUSION: $L^*(\hat{\eta}|\underline{x}) = L^*[\tau(\hat{\theta})|\underline{x}]$, $\tau(\hat{\theta})$ IS AN MLE FOR $\tau(\theta)$.

EX: FOR THE NORMAL $(\mu, 1)$ MODEL THE MLE FOR μ^2 IS $\hat{\mu}^2 = \bar{X}^2$.

FOR THE BERNOULLI MODEL THE MLE OF $\theta^{1/3}(1-\theta)^{1/2}$ IS $\bar{X}^{1/3}(1-\bar{X})^{1/2}$.

METHOD OF MOMENTS.

IDEA MATCH K-SAMPLE MOMENTS WITH K-POPULATION MOMENTS AND SOLVE FOR θ

LET $X_1, X_2, \ldots X_n$ BE A ~~pdf~~ SAMPLE WITH pdf OR pmf $f(\underline{x}|\theta_1, \theta_2 \ldots \theta_k)$, $\underline{\theta} = (\theta_1, \theta_2 \ldots \theta_k)$

DEFINE

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i^1, \quad \mu_1' = E(X^1)$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2; \quad \mu_2' = E(X^2)$$

$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k; \quad \mu_k' = E(X^k)$$

NOTE THAT EACH POPULATION MOMENT μ_j' IS A FUNCTION OF Θ.
i.e. $\mu_j' = \mu_j'(\theta_1, \theta_2 \ldots \theta_K)$

THE METHODS OF MOMENT ESTIMATOR $(\tilde{\theta})$ IS OBTAINED BY SOLVING THE EQUATIONS FOR $(\theta_1, \theta_2 \ldots \theta_K)$ IN TERMS OF $(m_1, m_2 \ldots m_K)$.

$$m_1 = \mu_1'(\theta_1, \theta_2 \ldots \theta_K)$$
$$m_2 = \mu_2'(\theta_1, \theta_2 \ldots \theta_K)$$
$$\vdots \qquad\qquad \vdots$$
$$m_K = \mu_K'(\theta_1, \theta_2 \ldots \theta_K)$$

EX. $X_1, X_2 \ldots X_n$ iid $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$

$\mu_1' = E(X) = \mu$; $\mu_2' = E(X^2) = \mu^2 + \sigma^2$

EQS.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (1) \quad \Rightarrow \hat{\mu} = \bar{x}$$
$$\mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \quad (2)$$

(2) $\Rightarrow$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) =$$

$$1/n\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$$

EX: $X_1, X_2 \ldots X_n$ iid $X_i \sim U(0, \theta)$

METHODS OF MOMENT ESTIMATOR FOR θ

$$f(x|\theta) = \frac{1}{\theta} I_{(0,\theta)}(x) \qquad E(X) = \int_0^\theta x \frac{1}{\theta}\,dx = \frac{1}{\theta}\frac{x^2}{2}\Big|_0^\theta = \frac{\theta}{2}$$

$$\Rightarrow \sum_{i=1}^{n} x_i / n = \theta/2 \quad \Rightarrow \hat{\theta} = 2\bar{x}$$

(DIFFERENT TO MLE $\hat{\theta} = \max\{X_1, X_2 \ldots X_n\}$)

$X_1, X_2 \ldots X_n$ iid observations such that $X \sim N(0, \sigma^2)$

$E(X) = 0$, $m_1 = \bar{X} \Rightarrow \bar{X} = 0$?

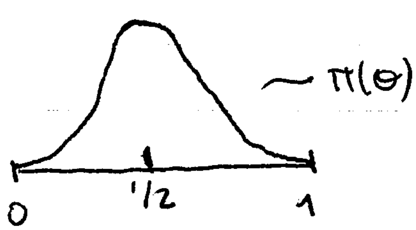
$E(X^2) = \sigma^2$, $m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

IF $X \sim \text{CAUCHY}(\theta) \Rightarrow f(x|\theta) = \left(\frac{1}{\pi}\right)\frac{1}{1+(x-\theta)^2}$

MEAN AND VARIANCE DO NOT EXIST
mgf DOES NOT EXIST $\Rightarrow$ THE METHOD OF MOMENTS ESTIMATOR DOES NOT EXIST. (REST IN PARAMETER SPACE)

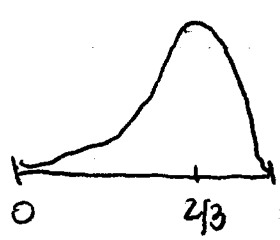
BAYES ESTIMATORS:

IN THE BAYESIAN APPROACH, WE CONSIDER θ TO BE A RANDOM QUANTITY SUBJECT TO PROBABILITY ASSESSMENT (THE PRIOR DISTRIBUTION) $\pi(\theta)$
THE PRIOR REFLECTS THE SUBJECTIVE KNOWLEDGE THAT A RESEARCHER HAS ABOUT θ AND IS NOT BASED ON DATA, BUT ON PREVIOUS EXPERIENCE OR HISTORY.
FOR EX: IN THE PROBLEM OF ESTIMATING θ IN A BERNOULLI MODEL A POSSIBLE PRIOR COULD BE A pdf LIKE:



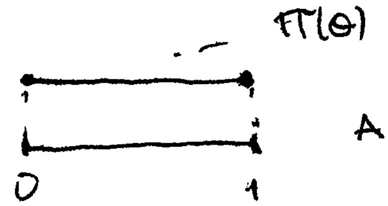
SAYING THAT θ IS AROUND $1/2$

FOR ANOTHER EXPERIMENTER: THE PRIOR COULD BE



FAVORING θ AROUND $2/3$.

SOMEONE THAT DOES NOT KNOW MUCH ABOUT θ COULD USE:



A $U(0,1)$

ANYWAY, THEY'RE ALL VALID STATEMENTS SINCE PRIORS ARE SUBJECTIVE.

WHEN THE DATA IS COLLECTED $\underline{X} = (X_1, X_2, \ldots X_n)$ WE UPDATE THE INFORMATION OF THE PRIOR $\pi(\theta)$ WITH BAYES THEOREM TO OBTAIN THE POSTERIOR DISTRIBUTION $\pi(\theta|\underline{x})$

IN FACT, BAYES THEOREM ESTABLISHES:

$$\pi(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)\,\pi(\theta)}{m(\underline{x})}; \quad m(x) = \int f(x|\theta)\,\pi(\theta)\,d\theta$$

WHERE $m(\underline{x})$ IS THE MARGINAL DISTRIBUTION OF $\underline{X}$

A BAYES ESTIMATOR IS A SUMMARY OF $\pi(\theta|\underline{x})$, A MEDIAN OR A MEAN OF THIS DISTRIBUTION.

EX: $X_1, X_2, \ldots X_n$ BE iid BERNOULLI (θ) RVs.

SUPPOSE THE PRIOR DIST. ON θ IS A BETA (α, β) α, β FIXED KNOWN

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad 0<\theta<1$$

$$f(\underline{x}|\theta) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

$$m(\underline{x}) = \int_0^1 \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}\,\pi(\theta)\,d\theta = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 \theta^{\alpha+\Sigma x_i-1}(1-\theta)^{\beta+n-\Sigma x_i-1}\,d\theta$$

INSIDE THE INTEGRAL WE RECOGNIZE THE FORM OF A BETA, pdf.

$$\Rightarrow \int_0^1 \theta^{\alpha+\sum x_i - 1}(1-\theta)^{\beta+n-\sum x_i - 1}\, d\theta = \frac{\Gamma(\alpha+\sum x_i)\,\Gamma(\beta+n-\sum x_i)}{\Gamma(\alpha+\beta+n)} \times$$

$$\int_0^1 \text{BETA}\,(\alpha + \textstyle\sum x_i,\ \beta+n-\sum x_i) \overset{=1}{} \quad \Rightarrow \quad \alpha' = \alpha + \sum_{i=1}^{n} x_i, \quad \beta' = \beta + n - \sum_{i=1}^{n} x_i$$

$$m(\underline{x}) = \frac{\Gamma(\alpha+\sum x_i)\,\Gamma(\beta+n-\sum x_i)}{\Gamma(\alpha+\beta+n)} \times \Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta) = \frac{\Gamma(\alpha')\Gamma(\beta')}{\Gamma(\alpha'+\beta')} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

$$\Rightarrow \pi(\theta|\underline{x}) = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\frac{\Gamma(\alpha')\Gamma(\beta')}{\Gamma(\alpha'+\beta')}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}$$

$$\Rightarrow \pi(\theta|\underline{x}) = \frac{\Gamma(\alpha'+\beta')}{\Gamma(\alpha')\Gamma(\beta')}\,\theta^{\alpha'-1}(1-\theta)^{\beta'-1} \equiv \text{BETA}(\alpha', \beta')$$

A BAYES POINT ESTIMATOR COULD BE THE MEAN OF THIS BETA

$$\hat{\theta}_B = \frac{\alpha'}{(\alpha'+\beta')} = \frac{\alpha + \sum_{i=1}^{n} x_i}{(\alpha+\beta+n)}$$

(FUNCTION OF α, β AND THE DATA $\underline{X}$)

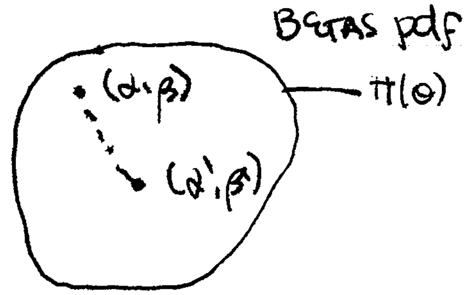
NOTICE THAT:

$$\hat{\theta}_B = \underbrace{\left(\frac{\alpha+\beta}{(\alpha+\beta+n)}\right)}_{C1}\underbrace{\left(\frac{\alpha}{\alpha+\beta}\right)}_{\text{PRIOR MEAN}} + \underbrace{\left(\frac{n}{\alpha+\beta+n}\right)}_{C2}\underbrace{\bar{X}}_{\text{SAMPLE MEAN}}$$

"POOLING" INFORMATION FROM THE PRIOR AND THE DATA

NOTICE THAT WE STARTED WITH A BETA PRIOR AND FINISHED WITH A BETA POSTERIOR

THEN, THE BETA DISTRIBUTION IS CONJUGATE FOR BERNOULLI MODEL



WHERE $\alpha' = \alpha + \sum_{i=1}^{n} x_i$; $\beta' = \beta + n - \sum_{i=1}^{n} x_i$ (SIMPLE UPDATING FOR THE PARAMETERS)

SUPPOSE $X_1, X_2, \ldots X_n$ ARE iid $Exp(\theta)$, WHERE $\theta = 1/\lambda$

$\Rightarrow f(x|\lambda) = \lambda e^{-\lambda x}$, $\lambda > 0$, $x > 0$

FIND A CONJUGATE PRIOR FOR λ.

DETERMINED BY THE LIKELIHOOD FUNCTION; ...

$$f(\underline{x}|\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$ (SUFFICIENT STATISTIC $\Rightarrow$ CONJUGATE PRIOR)

NOTICE THAT AS A FUNCTION OF λ, THIS LOOKS LIKE A GAMMA pdf WITH $(\alpha = n+1 ; \beta = 1/\sum x_i)$

SO GAMMA SHOULD WORK AS A CONJUGATE MODEL OR FAMILY.

$$\pi(\lambda) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda/\beta}, \quad \lambda > 0, \ \alpha > 0, \ \beta > 0$$

(α, β FIXED BY THE EXPERIMENTER)

$$m(\underline{x}) = \int_0^\infty \lambda^n e^{-\sum_{i=1}^n x_i(\lambda)} \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta} d\lambda$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty \lambda^{\alpha+n-1} e^{-(\sum x_i + 1/\beta)\lambda} d\lambda \text{ ; MAKE } \begin{array}{l} \alpha' = \alpha+n \\ \beta' = (1/\beta + \sum x_i)^{-1} \end{array}$$

$$= \frac{\Gamma(\alpha')(\beta')^{\alpha'}}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty \text{GAMMA}(\alpha', \beta')\, d\lambda = \frac{(\beta')^{\alpha'}}{\beta^\alpha} \frac{\Gamma(\alpha')}{\Gamma(\alpha)}$$

POSTERIOR

$$\pi(\lambda|\underline{x}) = \frac{\lambda^n e^{-\lambda \sum_{i=1}^n x_i} \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta}}{(\beta')^{\alpha'} \Gamma(\alpha') / \beta^\alpha \Gamma(\alpha)} = \frac{1}{(\beta')^{\alpha'}\Gamma(\alpha')} \lambda^{\alpha'-1} e^{-\lambda/\beta'}$$

A BAYES ESTIMATOR COULD BE THE MEAN OF λ WITH RESPECT TO THE POSTERIOR (REMEMBER THE MEAN OF GAMMA $= \alpha\beta$)

$$\hat{\lambda}_B = \alpha'\beta' = (\alpha+\beta)\left(1/\beta + \sum_{i=1}^n x_i\right)^{-1}$$

COMMENTS ABOUT BAYES:

i) RESULTS ~~ARE~~ MAY CAN BE PRIOR, DEPENDENT. → ADVANTAGE OF BAYESIAN → THE MOST CONTROVERSIAL POINT

ii) CALCULATIONS OUTSIDE CONJUGATE FAMILIES MAY BE HARD AND MAY REQUIRE NUMERICAL INTEGRATION FOR $m(\underline{x})$ (NOT GOING TO COVER THIS FOR THIS CLASS)

SKIP SECTION 7.2.4 (EM ALGORITHM)

FRIDAY START OF SECTION 7.3 . EVALUATION OF ESTIMATORS