Section 4: Bayesian Statistical Inference

- *Purpose*
  - *Students will learn subjectivist view of probability, use of Bayesian updating, and applications to commonly encountered kinds of data*
- *Objectives*
  - *Students will learn*
    - *Probability interpreted as a quantification of degree of plausibility*
    - *Bayes' Theorem, Bayesian updates*
    - *Use of discrete priors*
    - *Conjugate priors for Poisson, binomial, and exponential data*
    - *Model validation, checking consistency of data and prior*
    - *Jeffreys noninformative prior for Poisson, binomial, and exponential data*
    - *Techniques for using other priors such as lognormal*

Bayesian Statistical Inference

- *Topics to be covered*
  - *Logic of Probability Theory*
    - *Subjective probability*
    - *Bayes' Theorem*
  - *Prior distributions*
    - *Discrete and continuous*
    - *Conjugate*
    - *Noninformative*
    - *Nonconjugate*

Bayesian Statistical Inference

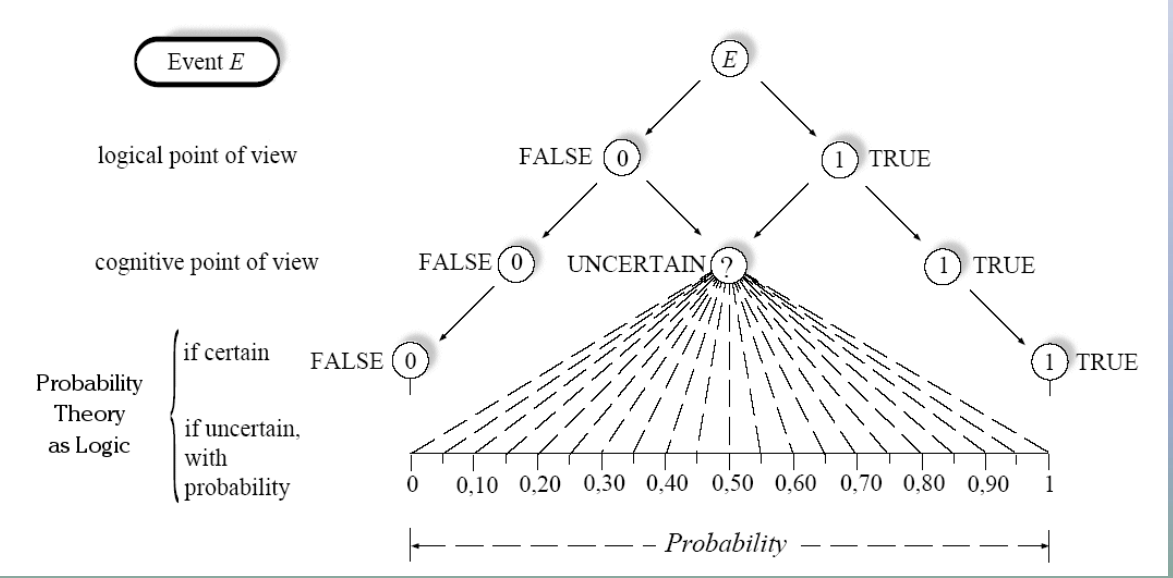
- General framework is covered in HOPE…
  - Page 6-2 (one-page introduction)
  - Section 6.2.2 for initiating events and running failures
    - Failure to run is also covered in Section 6.5
  - Section 6.3.2 for failures on demand
  - Section 8.3 for hierarchical Bayes methods
  - Section B.5 for summary of Bayesian estimation

Motivation for Bayesian Inference

- *Two problems with frequentist inference*
  - *If data are sparse, MLE can be unrealistic*
  - *No way to propagate uncertainties through the logic model*
- *Solution: A different interpretation of "probability"*
  - *Information about the parameter, beyond what is in the data, is included in the estimate*
  - *Use simulation to pass the uncertainties through the logic model*

Bayesian Statistical Inference

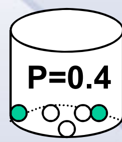
- *In the **Bayesian**, or "subjectivist," approach, probability is a **quantification of degree of belief***
  - *It is used to describe the plausibility of an event*
    - *Plausibility – "Apparent validity"*
  - *A mechanism to encode **information***
- *Note that, for "Bayes' Theorem,"*
  - *Thomas Bayes never wrote it*
  - *Laplace first used it in real problems*



Bayesian Statistical Inference (cont.)

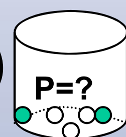
- *So what we have is three types of probability*

Classical (since 1600's)



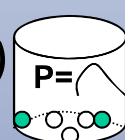
Equally likely outcomes, P(green) = 0.4 Known sample space, no data, thus no confidence interval

Frequentist (since 1920's)



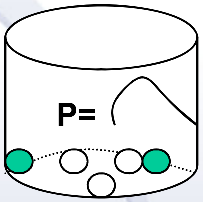
P=lim x/n as n $\rightarrow\infty$, data used for inference, confidence intervals used

Bayesian (1700's, 1930's-)



P depends only on state of knowledge, unlike classical, no such thing as an inherent P, we ***assign*** *probability*

Bayesian Statistical Inference (cont.)



- *Looking at the urn for the Bayesian case*
  - *We have different possible perspectives*

    P(p=# | completely ignorant) = ?

    P(p=# | knowledge of other urns, selected a couple from urn) = ?

    P(p=# | "cheated" by looking in urn) = ?
  - *Also, for Bayesian analysis, no "population" is needed*
    - *But of course, the more knowledge we have, the better off we are*
  - *Nuclear power plant PRA relies a great deal on the body of engineering knowledge in order to perform calculations*

Bayesian Parameter† Estimation

- *The general procedure is:*
  1. *Begin with a prior distribution about parameter, quantifying uncertainty, i.e., quantifying degree of belief about possible parameter values*
  2. *Observe and analyze data*
  3. *Obtain the posterior distribution for the parameter*
- *We follow this process to determine the probability that a hypothesis is true, conditional on* ***all*** *available evidence*
  - *This approach is fundamentally different from the classical statistics methods*

† Note that in PRA, other uncertainties exist, such as model uncertainty

Bayesian Parameter Estimation (cont.)

- *Consider the unknown parameter λ. Same idea if the parameter is p.*
- *For now, assume X is discrete, with $f(x|\lambda) = Pr(X=x|\lambda)$.*
- *Also assume that the unknown parameter λ can only take discrete values, λ_1, λ_2, …*
- *Define prior distribution, $g_{prior}(\lambda_i) = Pr(\lambda = \lambda_i)$.*
- *By Bayes' Theorem,*

$$\Pr(\lambda=\lambda_i)=\frac{\Pr(X=x\,|\,\lambda=\lambda_i)\Pr(\lambda=\lambda_i)}{\sum\Pr(X=x\,|\,\lambda=\lambda_j)\Pr(\lambda=\lambda_j)}$$

or

$$g_{posterior}(\lambda_i)=\frac{f(x\,|\,\lambda_i)g_{prior}(\lambda_i)}{\sum f(x\,|\,\lambda_j)g_{prior}(\lambda_j)}$$

- *Denominator is a normalizing constant.*

Bayesian Parameter Estimation (cont.)

- General version, including discrete and continuous cases
- Define $g_{prior}(\lambda)$, the prior p.d.f. of λ
- Let $f(x|\lambda)$ be the p.d.f. of X, dependent on λ
  - This is the **likelihood**
- The posterior p.d.f. of λ is

$$g_{posterior}(\lambda) \propto f(x \mid \lambda) g_{prior}(\lambda)$$

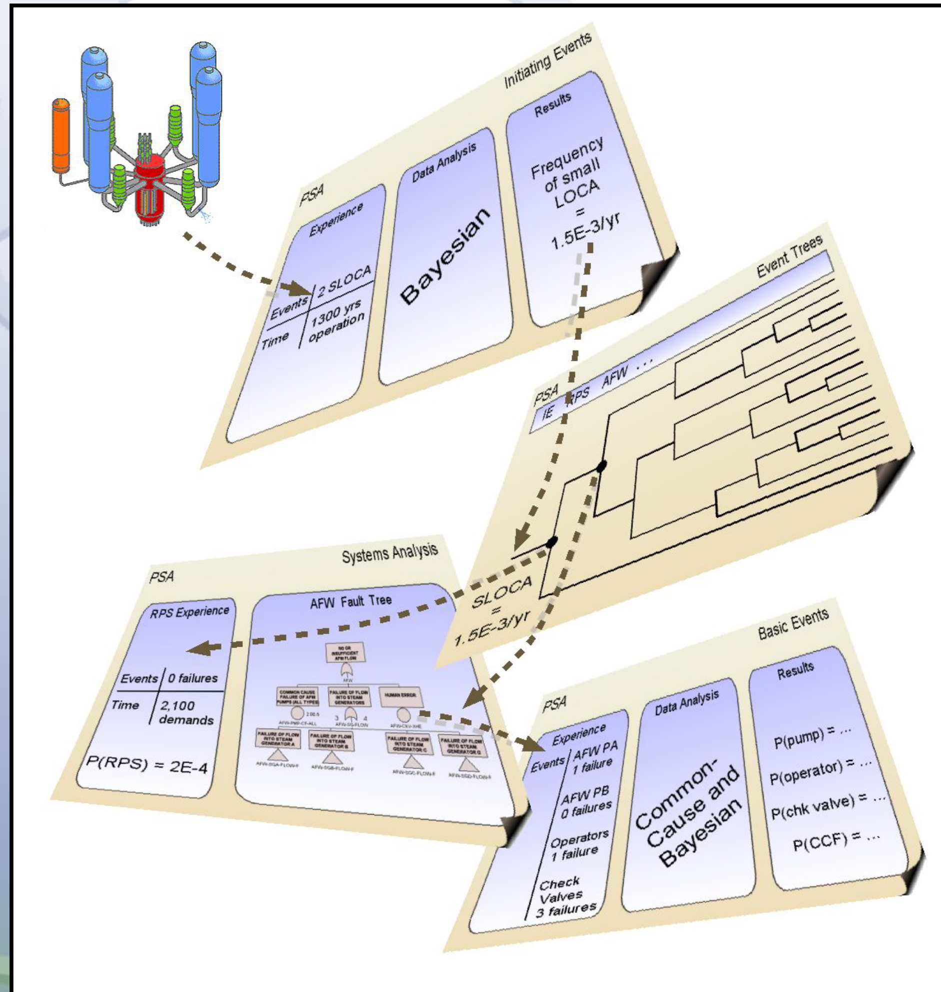
- $g_{posterior}$ **is proportional to** the expression on the right.

Bayesian Parameter Estimation (cont.)

- *If prior distribution is continuous, parameter has values over a continuous range (and infinitely many possible values).*
- *Even though our **goal** is to obtain a distribution [$g_{post}(\lambda \mid x)$] on a parameter λ, need to remember*
  - *λ is assigned a distribution (i.e., information)*
    - *Often convenient to summarize distribution by metrics such as mean, variance, or percentiles*
  - *Note that the distribution is subjective, not a real, physical distribution*
    - *For example, the "urn" does not change even though information (and probability) on its contents may change*

Bayesian Parameter Estimation (cont.)

- *In PRA, we repeat the estimation process and incorporate into our models*



Uses of Posterior Distribution

- *For presentation purposes*
  - *Plot the posterior p.d.f.*
  - *Give the posterior mean*
  - *Give a **Bayes credible interval**, an interval that contains most of the posterior probability (e.g. 90% or 95%)*
- *For risk assessment*
  - *Sample from the distribution of each parameter*
  - *Combine the results to obtain sample from Bayes distribution of end-state frequency*

Historical Use of Bayes Theorem

- *Laplace, in 1774, used Bayesian methods to estimate the mass of Saturn*
  - *Assumed uniform prior density (what was known at the time)*
  - *Data consisted of mutual perturbations between Jupiter and Saturn*
- *His result was that he gave* ***odds*** *of 11,000 to 1 that his mass estimate is not in error by more than 1%*
  - *What do odds of 11,000 to 1 imply?*
    - *That the estimate ± 1% is 99.99% credible interval*
- *200 years of science increased his estimate by about 0.6%*
  - *Laplace would have won his bet (so far!)*

PRA Applications of Bayes' Theorem

- *We are going to focus on three different situations related to **different types** of prior information*
  - *Discrete priors*
  - *Conjugate priors*
    - *Informative*
    - *Noninformative*
- *We will discuss, but not focus on, nonconjugate priors*

Discrete Prior Distributions

- *These priors are easy with a spreadsheet*
  - *Follows directly from Bayes' Theorem*
    - *For example, see page 4-7*

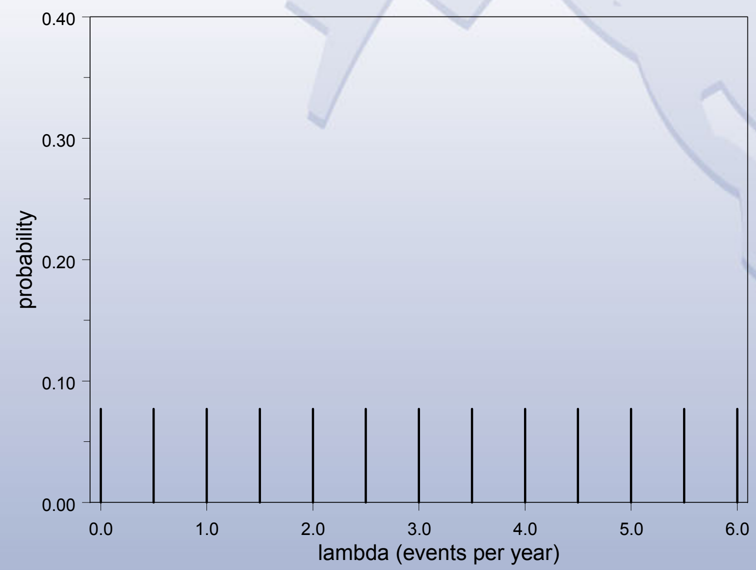
$$\Pr(\lambda = \lambda_i) = \frac{\Pr(X = x \mid \lambda = \lambda_i)\Pr(\lambda = \lambda_i)}{\sum \Pr(X = x \mid \lambda = \lambda_j)\Pr(\lambda = \lambda_j)}$$

- *Numerator on right in Bayes' Theorem is the product of the likelihood times the prior probability of λ_i*
  - *To obtain full posterior probability, divide every such product by the sum of all such products*
  - *This makes the posterior probabilities (for all possible values of λ_i) sum to 1.0*

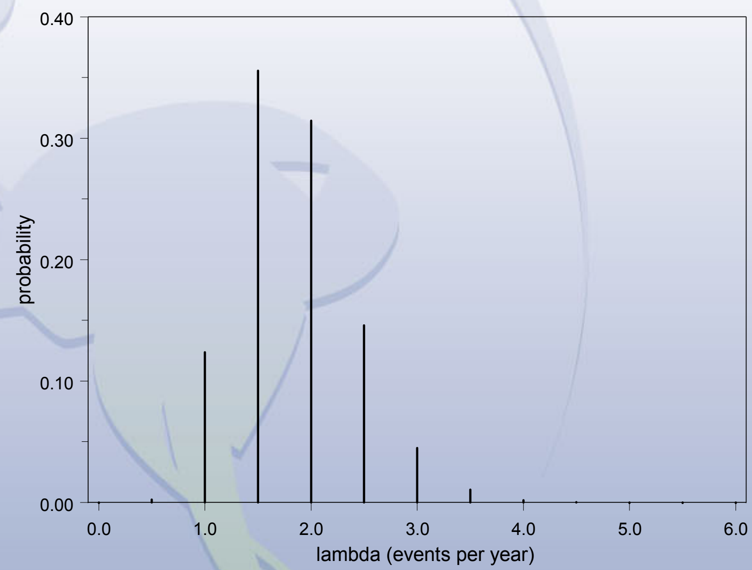
HOPE

Pages 6-9 through 6-11, 6-33 through 6-35

Example of Discrete Prior and Posterior



Coarse discrete prior for λ (events per year)



Posterior for λ, based on 10 observed events in 6 years

HOPE

6-10

Conjugate Priors

- *In this section, we will address **three** common cases found in PRA*
  - *Poisson process*
  - *Exponential process*
  - *Binomial process*

Industry Priors for LOSP Example (For Later Reference)

- λ_{LOSP}~gamma(13.8, 747 reactor-critical years)
  - From "Evaluation of Loss of Offsite Power Events at Nuclear Power Plants: 1986-2003," NUREG/CR-????, 2005?
  - Above result is built from several subtypes of LOSP event
  - Above result excludes Aug. 14, 2003 widespread blackout (simultaneous LOSP at 8 operating plants)
- p_{FTS}~beta(0.957, 190)
  - From Eide (2003)
- λ_{FTR}~gamma(1.32,1137 hrs)
  - From Eide (2003)
  - Above result constructed from two rates in paper

Conjugate Prior – Poisson Data

- *Facts about* **gamma(α, β)** *distribution, see HOPE*
  - *gamma(α, β) density is*
    - $g(\lambda) = C\,\lambda^{\alpha-1}e^{-\lambda\beta}$
      - *when $\alpha = 1$, then you have the* ***exponential*** *distribution*
  - *mean = α / β*
    - *Example, if $\alpha = 1$ and $\beta = 10$ then mean = 0.1*
  - *variance = α / β^2*
  - *100p percentile = $[\chi^2_p(2\alpha)] / (2\beta)$*

HOPE

Pages A-18 through A-20

Conjugate Prior – Poisson Data (cont.)

- Some programs, including Excel, use 1/ β instead of β as the second parameter for a **Gamma** distribution
  - Example, $\alpha=1.5$, β=10, in Excel

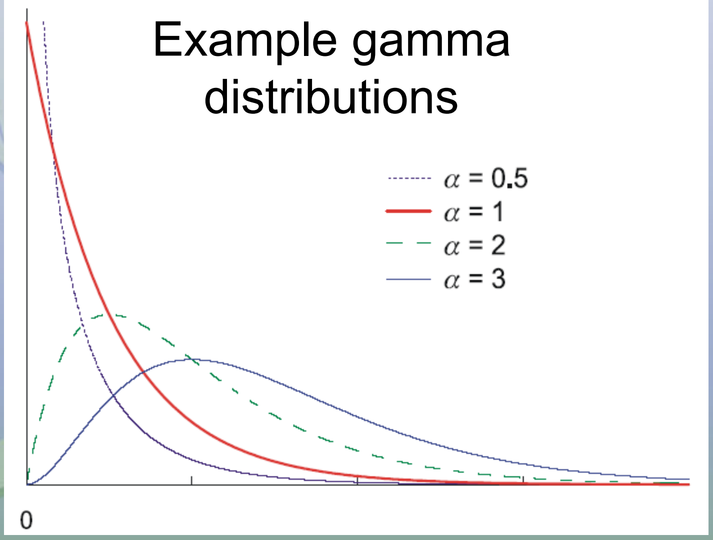
    =Gammainv(p, α, 1/ β)

    =Gammainv(0.05, 1.5,0.1) = 0.0176 (which is the 5%)
  - Using chi-square (on previous page)

    $= [\chi^2_p(2\alpha)] / (2\beta)$

    $= [\chi^2_p(3)] / (20)$

    = 0.35/20 = 0.018
  - In SAPHIRE, specify mean and alpha, which gives 5th=0.018



HOPE

Pages A-18 through A-20

Conjugate Prior – Poisson Data (cont.)

- *Now, we have a component with* ***x*** *failures in time* ***t***
- *If X is* **Poisson**(λt) *and* $g_{prior}(\lambda)$ *is* **gamma**$(\alpha_{prior}, \beta_{prior})$
  - *Then posterior distribution of λ is*
    - **gamma**$(\alpha_{posterior}, \beta_{posterior})$

      $\alpha_{posterior} = \alpha_{prior} + x$ (**x** = # events)

      $\beta_{posterior} = \beta_{prior} + t$ (*t = time related to seeing* ***x*** *events*)
    - *Therefore, posterior mean =* $(\alpha_{prior} + x)/(\beta_{prior} + t)$
      - *compromise between MLE, x/t*

        and prior mean, $\alpha_{prior} / \beta_{prior}$
    - *Posterior intervals generally shorter than those from data alone or prior alone*

HOPE **Pages 6-12, 6-13**

Conjugate Prior – Poisson Data (cont.)

- *If X is* Poisson*(λt) and $g_{prior}(\lambda)$ is* gamma*(α_{prior} , β_{prior}), we indicated that $g_{posterior}(\lambda)$ is* **also** *a gamma distribution*

  $\alpha_{posterior} = \alpha_{prior} + x$ (**x** = # events)

  $\beta_{posterior} = \beta_{prior} + t$ (*t = time related to seeing* **x** *events)*

- *Thus, the α parameter is* ***like*** *the number of failures while the β parameter is* ***like*** *the time period*
  - *If a PRA or a database indicates $\alpha = 1$ and β = 100 hr, then this is like seeing one failure in 100 hours of operation*

HOPE **Pages 6-12, 6-13**

Conjugate Prior – Exponential Data

- *If $T_1,\ldots,T_n$ are independent* **exponential***(λ) and $g_{prior}(\lambda)$ is* **gamma***(α_{prior}, β_{prior})*
  - *Then posterior distribution of λ is*
    - **gamma***($\alpha_{posterior}$, $\beta_{posterior}$)*

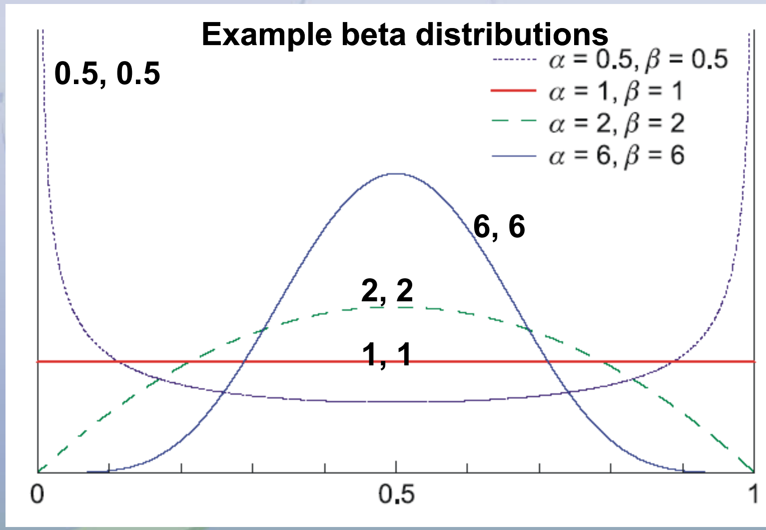
      $\alpha_{posterior} = \alpha_{prior} + n$ *(n = # events)*

      $\beta_{posterior} = \beta_{prior} + \Sigma t_i$ *(t_i = times related to seeing n events)*
- *Again, for a* **gamma(α, β)** *distribution*
  - *mean = α / β*
  - *variance = α / β^2*
  - *100p percentile = $[\chi^2_p(2\alpha)] / (2\beta)$*

HOPE . 6-61

Conjugate Prior – Binomial Data

- *Facts about* **beta(α, β)** *distribution*
  - *beta(α, β) density is g(p) = C $p^{\alpha-1}(1-p)^{\beta-1}$*
  - *mean = α / (α + β)*
  - *variance = mean(1 – mean)/(α + β + 1)*
  - *Tables in Handbook, App. C*
    - *%-tiles in Table C.4*
  - *For more accuracy, use BETAINV in Excel =Betainv(p, α, β)*



HOPE

Pages A-21 through A-22, C-8 through C-13

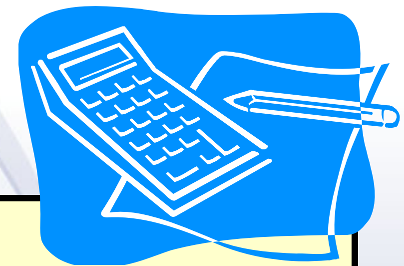
Conjugate Prior – Binomial Data (cont.)

- *Now, we have a component with **x** failures in **n** demands*
- *If X is* **binomial***(n, p) and* $g_{prior}(p)$ *is* **beta**$(\alpha_{prior}, \beta_{prior})$
  - *Then posterior distribution of p is*
    - **beta**$(\alpha_{posterior}, \beta_{posterior})$
      - $\alpha_{posterior} = \alpha_{prior} + x$ *(x = # events)*
      - $\beta_{posterior} = \beta_{prior} + n - x$ *(n = total # trials)*
    - *Posterior mean is* $(\alpha_{prior} + x)/(\alpha_{prior} + \beta_{prior} + n)$
    - *Again, this is a compromise between MLE, x/n and prior mean,* $\alpha_{prior} /(\alpha_{prior} + \beta_{prior})$
    - *Also,* α *is* ***like*** *the # of failures and* β *is* ***like*** *# of successes*

HOPE

Pages 6-35 through 6-37

Summary of Bayesian Estimates for LOSP Example



Parameter	Distribution	Point Est. (Mean)	90% Interval
λ_{LOSP}	*Industry Prior*	*1.8E-2* yr^{-1}	*(1.1E-2, 2.7E-2)* yr^{-1}
	Posterior	*2.0E-2* yr^{-1}	*(1.2E-2, 2.9E-2)* yr^{-1}
p_{FTS}	*Industry Prior*	*5.0E-3*	*(2.3E-4, 1.5E-2)*
	Posterior	*7.4E-3*	*(1.3E-3, 1.8E-2)*
λ_{FTR}	*Industry Prior*	*1.2E-3* hr^{-1}	*(1.1E-4, 3.2E-3)* hr^{-1}
	Posterior	*1.0E-3* hr^{-1}	*(9.6E-5, 2.8E-3)* hr^{-1}

Compare with frequentist results on p. 3-35

Comparing Prior and Data

- *The data and the prior should be consistent*
  - *This comparison is an aspect of model validation*
- *Picture: draw likelihood and prior distribution, see if they largely overlap*
- *Test of hypothesis that data and prior are consistent: calculate the marginal distribution of the data, equal to*

$$\Pr(X = x) = \int \Pr(X = x \mid \lambda) g_{prior}(\lambda) d\lambda$$

  - *Here the parameter is λ. Same idea if parameter is p.*
- *See if observed x is in either tail of this distribution*
  - *Pr(X ≤ observed x) is small or*
  - *Pr(X ≥ observed x) is small.*
  - *If so, question whether prior and data are consistent*

HOPE

Pages 6-30, 6-50, 6-51

How to Get a Prior?

- *Four general types of priors*

Simplicity ↑

1. *Simple discrete (e.g., spreadsheet)*
2. *Conjugate (simple formulas, lookup tables)*
3. *Noninformative (some simple, some not simple)*
4. *General (numerical techniques via computers)*

- *Prior should reflect belief/knowledge*
  - *But, people are (generally) terrible "probability generators"*
- *Warnings in HOPE*
  - *Beware of zero values for prior*
  - *Avoid conservatism (PRA is best estimate)*
- *When data is plentiful, do not invest too much effort in determining the prior*

Noninformative Prior Distributions

"Ignorance is preferable to error…" (Thomas Jefferson, 1781)

- *The point of "**noninformative**" priors is to answer question*
  - *How do we find a prior representing complete ignorance?*
- *Rev. Bayes suggested a **uniform** prior*
  - *Laplace used this in his activities with great success*
  - *But, there are philosophical/mathematical problems with this*
- *Jeffreys **suggested** a prior that was "ignorant" to variations in*
  - *Scale*
  - *Location*

Noninformative Prior Distributions (cont.)

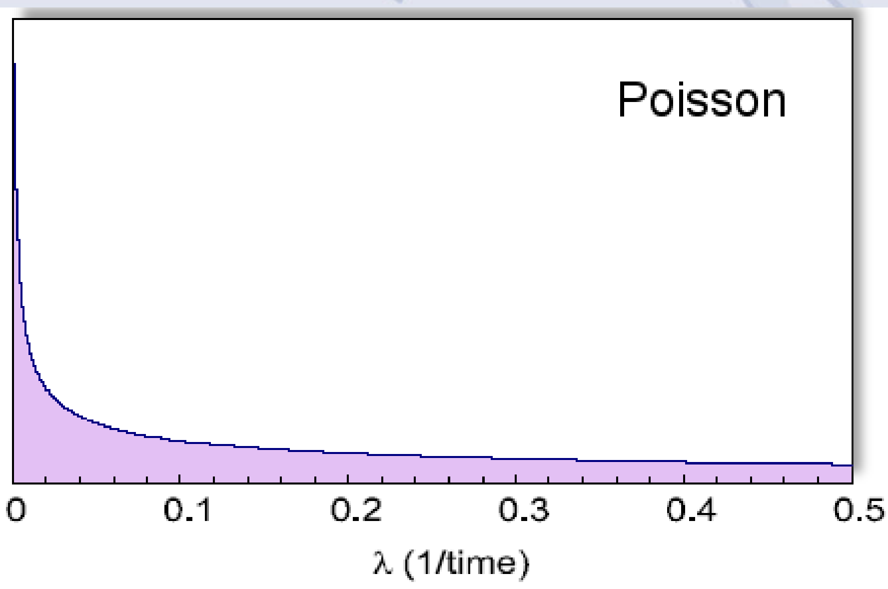
- *Consequently, so-called "noninformative" prior is typically* ***not*** *uniform!*
  - *Instead, it depends on the process generating the data*
  - *Has property that the Bayes posterior intervals are* ***approximately*** *equal to Frequentist confidence intervals*
- *For* **Poisson**(λt) *data*
  - *Noninformative prior for* λ *is proportional to* **gamma***(1/2, 0)*
- *For* **exponential**(λ) *data*
  - *Noninformative prior for* λ *is proportional to* **gamma***(0, 0)*
- *For* **binomial***(n, p) data*
  - *Noninformative prior for p is* **beta***(1/2, 1/2)*

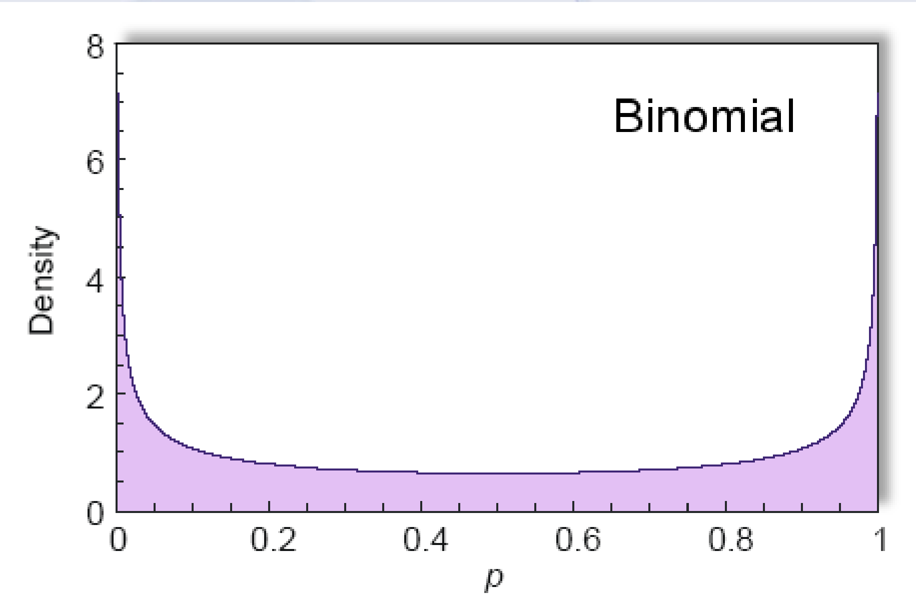
HOPE

Pages B-12 through B-14, 6-14, 6-37, 6-61, 6-62

Noninformative Prior Distributions (cont.)

- *Two of these priors look like*

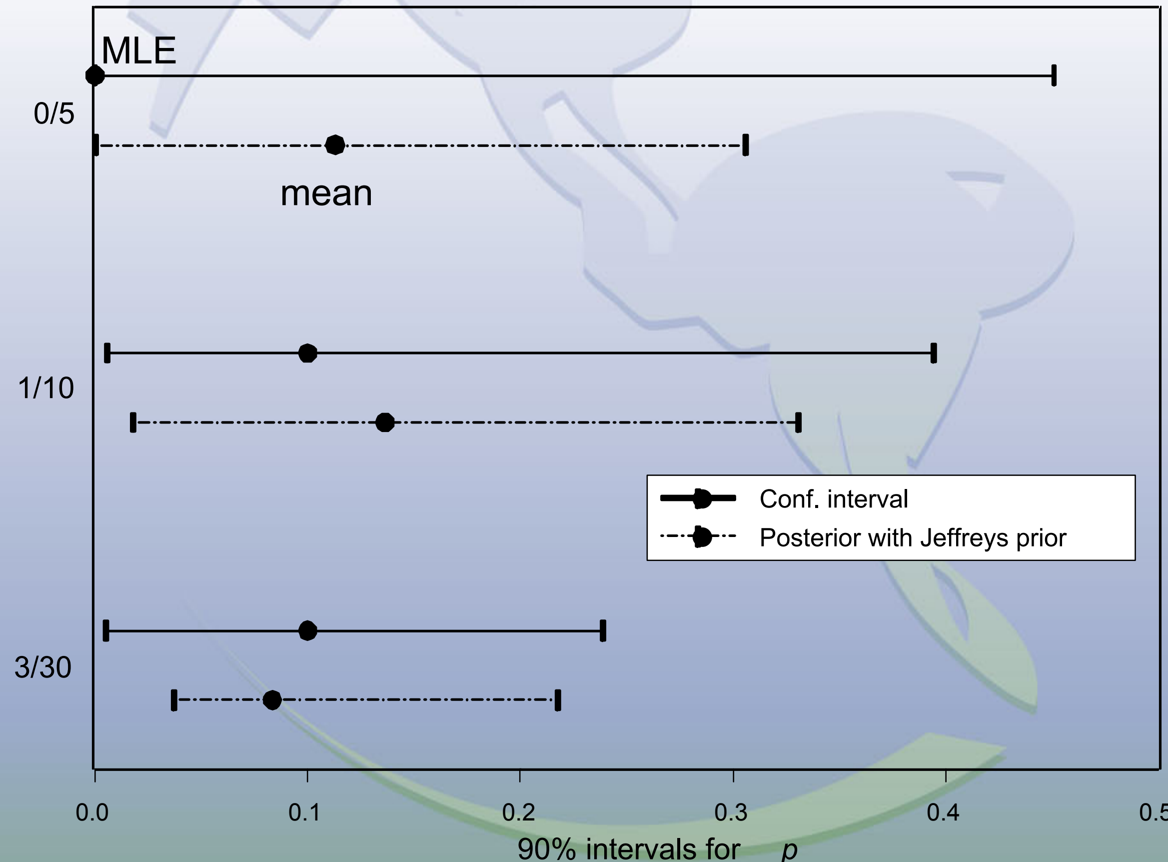




gamma(1/2, 0)　　　　*beta(1/2, 1/2)*

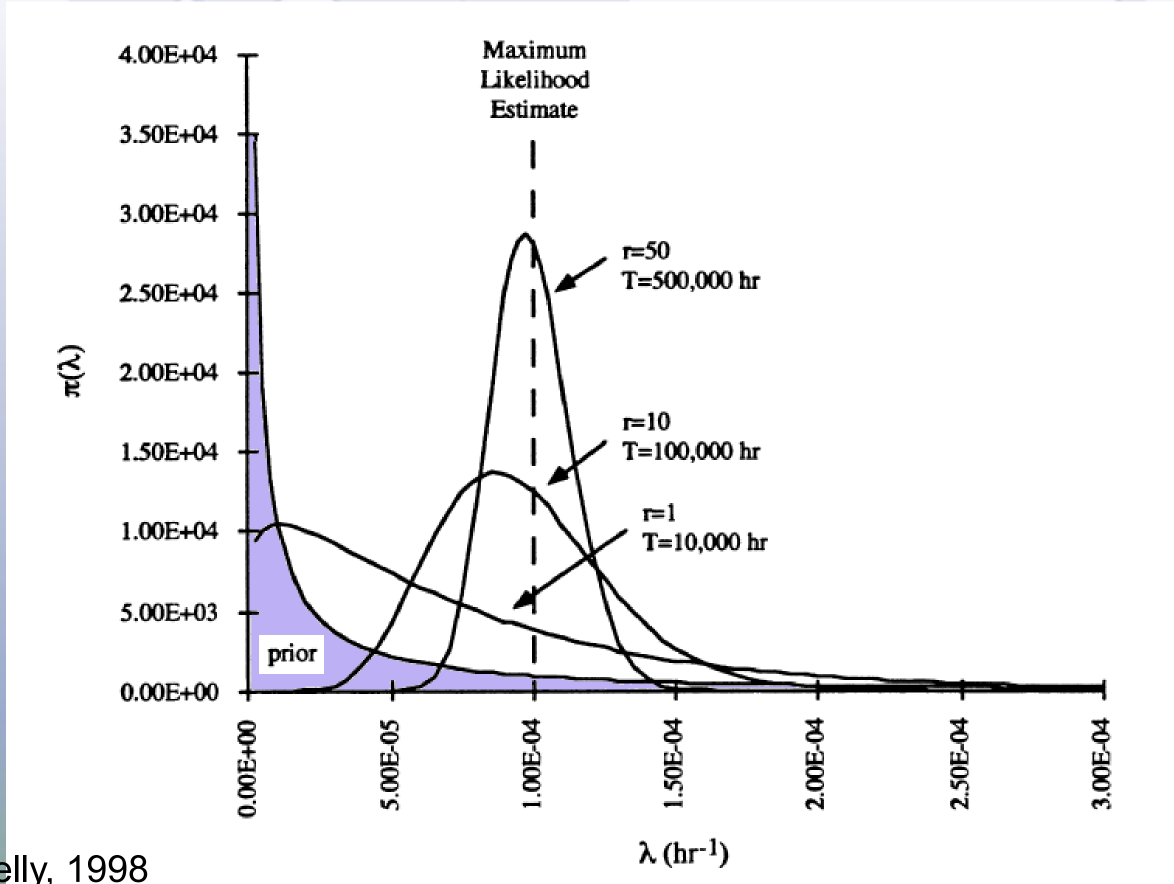
Noninformative Prior Distributions (cont.)

- *Let us compare results from these priors to confidence intervals*



Noninformative Prior Distributions (cont.)

- *But, as data increases, prior becomes less important*



Siu and Kelly, 1998

Lognormal Distribution

- *Definition of a lognormal distribution:*
  - *X is* **lognormal**$(\mu_{ln}, \sigma_{ln}^2)$ *if* $\ln(X)$ *is* **normal**(μ, σ^2)
- *Lognormal distribution is* ***often used*** *as a prior distribution in PRA, even though it is* ***not conjugate***
  - *Median of X is* e^{μ}
  - *Mean* (μ_{ln}) *of X is* $e^{\mu+(\sigma^2/2)}$
  - *Variance* (σ_{ln}^2) *of X is* $(\mu_{ln})^2\,(e^{\sigma^2}-1)$
  - *Error factor (EF) is* $e^{1.645\sigma}$
  - *Other ways to write EF (applies only to lognormal)*
    - $EF = 95^{th}/50^{th} = 50^{th}/5^{th} = (95^{th}/5^{th})^{1/2}$

HOPE

Pages A-16, A-17

Lognormal Distribution (cont.)

- $$\Pr(X \le x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$
  where Φ is tabulated in HOPE, Appendix C
- *Lognormal distribution is determined by* ***any two*** *of*
  - *μ*
  - *σ^2*
  - *mean (μ_{ln})*
  - *variance (σ_{ln}^2)*
  - *EF*

Known parameters				Equation
Mean (μ_{ln})	Standard deviation (σ_{ln})	Error factor (EF)	Median ($\tilde{x}_{ln}$)	
✓	✓			$EF = e^{1.645\sqrt{\ln(1+(\sigma_{ln}/\mu_{ln})^2)}}$
	✓		✓	$\mu_{ln} = e^{\ln(\tilde{x}_{ln}) + \frac{1}{2}\ln\left(0.5 + 0.5\sqrt{1+4(\sigma_{ln}/\tilde{x}_{ln})^2}\right)}$ $EF = e^{1.645\sqrt{\ln((\mu_{ln}/\tilde{x}_{ln})^2)}}$
		✓	✓	$\mu_{ln} = \tilde{x}_{ln}\,(EF)^{0.185\ln(EF)}$
✓			✓	$EF = e^{1.645\sqrt{\ln((\mu_{ln}/\tilde{x}_{ln})^2)}}$

Nonconjugate Prior Distributions

- *If posterior distribution is* **gamma**(α, β) *or* **beta**(α, β) *with* α *small (much smaller than 0.5), then lower tail of distribution is unrealistically large (5th percentile value very small)*
  - *Example, Gamma(0.125, 1 hr)*

    %-tile $= [\chi^2_p(2\alpha)] / (2\beta) = [\chi^2_{0.05}(0.25)] / 2 = 4.8E\text{-}11/2 \approx 2E\text{-}11/\text{hr}$
- *For this and other reasons, we may prefer a nonconjugate prior*
  - *When prior is not conjugate*
    - *Posterior distribution must be found by numerical integration or by simulation. This is mostly skipped in this course.*

HOPE

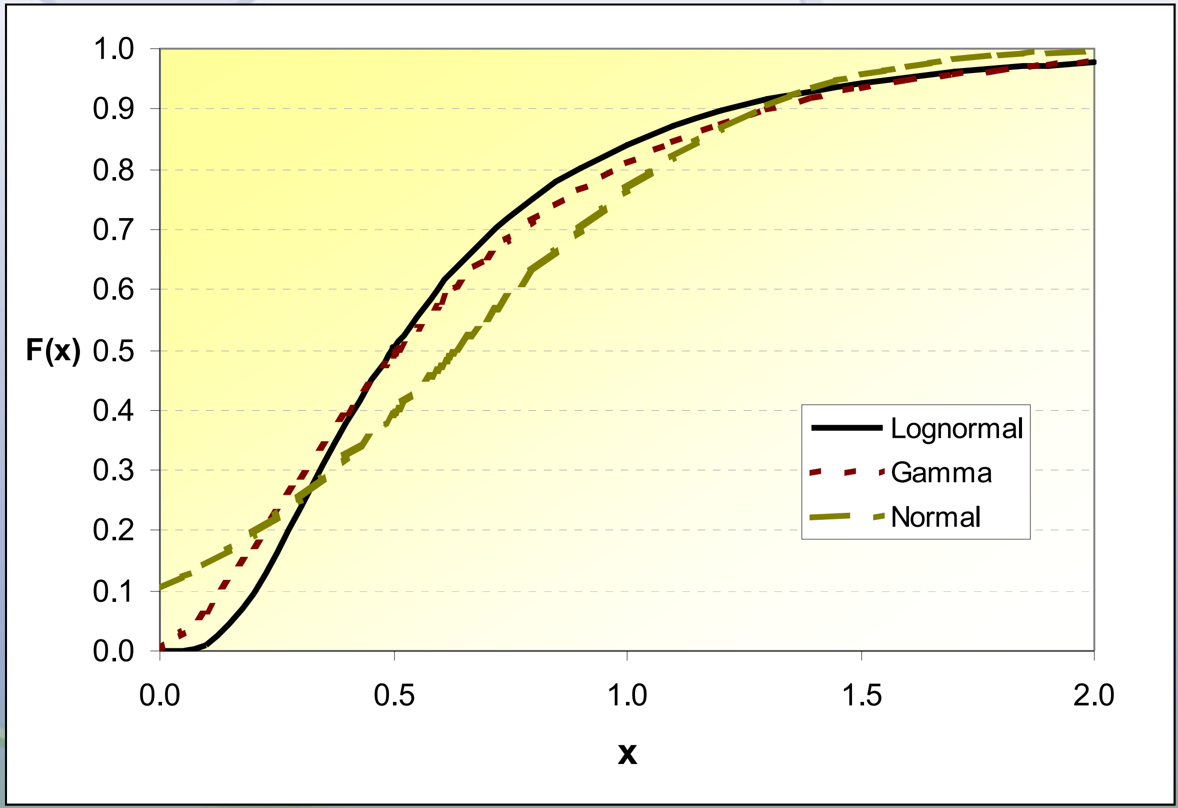
Pages 6-16 through 6-20, 6-39 through 6-43

Nonconjugate Prior Distributions (cont.)

- *Sometimes, we have a prior, such as a nonconjugate prior, and we may prefer **not** to use it*
- *We may select a "similar" prior of a **different**, but more convenient functional form (such as a conjugate prior)*
  - *To replace one prior by another, **easiest** way (in most cases) is to make them have same mean and same variance*
    - *This "**matching**" approach can be done with algebra alone*
  - *Other ways, such as making them have same mean and same 95th percentile, are typically harder*
    - *Note that "similar" priors do not necessarily have similar percentiles, or produce posterior distributions with similar percentiles*

Nonconjugate Prior Distributions (cont.)

- *For example, perhaps we have a **lognormal** prior*
  mean=0.63
  s.dev=0.50
- *We can find **other distributions** that have same moments*
  - *Gamma*
  - *Normal*



Monte Carlo Sampling

- *Approximate a distribution by generating a large random sample from the distribution*
- *Useful for*
  - *Propagating uncertainties through logic model (e.g. fault tree or event tree)*
  - *Approximating posterior distribution when it does not have simple form (e.g. when prior is not conjugate)*

Monte Carlo Sampling (cont.) Simulation of a Uniform(0,1) Distribution

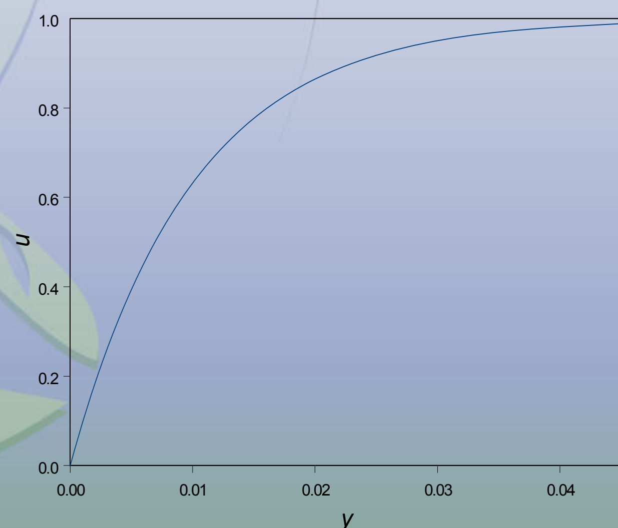
- *Many software packages can simulate uniform distribution*
  - *Excel, Visual Basic, FORTRAN, SAPHIRE, etc.*
- *Completely deterministic, not random*
  - *"Looks" random*
  - *Really, the output is a long (e.g. ~2^{31}) sequence of distinct numbers in unpredictable order*
  - *User inputs a "seed", or computer uses the clock time. This determines where in the sequence we start*

Monte Carlo Sampling (cont.) Simulation of a Binomial Random Variable

- *To simulate a* **binomial***(n,p) random variable, do this many times:*
  - *Generate n random numbers* u_1 *through* u_n *from a* **uniform***(0,1) distribution*
  - *If* $u_i < p$ *define* $x_i = 1$*. Otherwise define* $x_i = 0$*.*
  - *Set* $y = x_1 + ... + x_n$
- *The many values of y are a sample from a* **binomial***(n,p) distribution*

Monte Carlo Sampling (cont.) Use of "Inverse c.d.f. method"

- *Do the following many times*
  - *Generate u from a* **uniform***(0,1) distribution*
  - *Set* $y = F^{-1}(u)$*, where*
    - *F is the c.d.f. of Y,* $F(y) = Pr(Y < y)$
    - F^{-1} *is inverse function,* $F(y) = u \quad \Leftrightarrow \quad F^{-1}(u)=y$
- *Values of y are a random sample from the distribution of Y*
- *Idea…*
  - *Choose most values where F is steep*



Monte Carlo Sampling (cont.) Use of Transformation

- *For example, to generate lognormal Y*
  - *First generate n values from a normal distribution, with n large*
    - *Call them x_1 through x_n*
  - *Set $y_i = exp(x_i)$, so that $ln(y_i) = x_i$*
- *The y_i values are a random sample from a lognormal distribution*

Monte Carlo Sampling (cont.) How Big a Sample Is Needed?

- *Let true distribution of Y have mean μ and variance σ^2*
- *Generate (large) sample, y_1, …, y_n*
- *Estimate μ by sample mean, i.e. average of sample values, $\bar{y}$*
- *Approximate 95% confidence interval for μ is*

  $$\bar{y} \pm 2s / \sqrt{n}$$ *(n = # samples)*

  - *Here s is sample standard deviation, an estimate of σ*
  - *$s / \sqrt{n}$ is called the* ***standard error***

- *So to estimate μ and cut the error by a factor of 2, n must be increased by a factor of 4*

Extra Info – The Path to "Bayes' Theorem" (1 of 4)

- *Bayesian hypothesis tests leads to **Bayes' Theorem**, which leads us to the world of parameter estimation*
  - *This path will, ultimately, take us back to and allow us to solve problems such as LOSP*
- *As part of this path, **objectivity** requires that we take into account all evidence*
  - *Classical statistics use subset of evidence*
- ***One** of these parts of evidence is P(H | X)*
  - *H = hypothesis*
  - *X = general information known prior to having updated information (or data) specific to problem at hand*

The Path to "Bayes' Theorem" (2 of 4)

- *P(H | X) is the so-called* ***PRIOR***
- *A couple of comments on the "prior"*
  - *Prior information may come* ***later*** *in time than our updated information*
    - *For example, a scientist runs an experiment, but reads a journal article with surprising findings related to his work before he had a chance to analyze his data*
  - *One person's "prior probability" is another person's "posterior probability"*

The Path to "Bayes' Theorem" (3 of 4)

- *For inference on our hypotheses, we know the "data" and want to know, via probabilities, can our model give us those results*
  - *In other words, which hypothesis (from a set) is logically more plausible given the evidence and data available*
- *This **logical plausibility** between information implies*

$$P(D\,H \mid X) = P(D \mid H\,X)\,P(H \mid X)$$
$$= P(H \mid D\,X)\,P(D \mid X)$$

D = the data

H = our hypothesis

X = general information known prior to data

The Path to "Bayes' Theorem" (4 of 4)

- *We can take the product rule and use it since we desire*
  - *Probabilities **not** conditional on H*
  - *This point is, again, very much **unlike** classical statistics*
    - *Confidence intervals are **very much** conditional upon the hypothesis test being performed*
  - *It is this point that results in "Bayes' Theorem" occasionally being called the Theorem of Inverse Probabilities*
- *So, we directly write down, from the **product rule**, Bayes' Theorem*

$$P(H \mid D\,X) = P(H \mid X)\,\frac{P(D \mid H\,X)}{P(D \mid X)}$$