Upper Limits and Priors

James T. Linnemann

Michigan State University

FNAL CL Workshop

March 27, 2000

With notes added January, 2003

P(contents|I, finish)

prior probability or likelihood?

- Coverage of Cousins + Highand Limits
  - mixed Frequentist + Bayesian
- Dependence of Bayesian UL on
  - Signal "noninformative" priors
  - Efficiency informative priors
    - and comparison with C+H limits
  - background informative priors
- Summary and Op/Ed Pages

The Problem

- Observation: see **k** events
- Poisson variable:
  - expected mean is **s+b** (signal + background)
  - **s = $\varepsilon L \sigma$**
    - efficiency × Luminosity × cross section
      - "cross section " σ really cross section × branching ratio
- Calculate **U**, 95% upper limit on σ
  - function of k, b, and uncertainties δ_b, δ_ε, δ_L
  - focus on **upper limits: searches**

Some typical cases for Calculation of 95% Upper Limits

k=0, b=3 — The Karmen Problem

k=3, b=3 — Standard Model Rules Again

k=10, b=3 — The Levitation of Gordy Kane?

"seeing no excess, we proceed to set an upper limit…"

The 95% Solution:
Reverend Bayes to the Rescue

- Why? He appeals to our theoretical side

  from statistics, we want "the answer"; as close as it gets?

- Why? to handle nuisance parameters

Name your poison

- Tincture of Bayes

  Cousins and Highland treatment:

  - Frequentist signals + Bayesian nuisance

- Bayes Full Strength

  The DØ nostrum:

  Both signal and nuisance parameters Bayesian

Cousins & Highland

Trying to make everyone happy makes no one happy.
Not even Bob.

Treat **signal** in Frequentist fashion **(counts)**

Bayesian treatment of nuisance parameters

modifies probabilities entering signal distribution

"weighted average" over degree of belief in unknown parameters

Nota Bene

This is how every physicist I know instinctively approaches this problem. It's the "natural" way, particularly when writing a Monte Carlo

C+H Coverage Monte Carlo: b=0; sensitivity uncertainty

- Fix true sensitivity, σ in outer loop

  sweep through parameter space

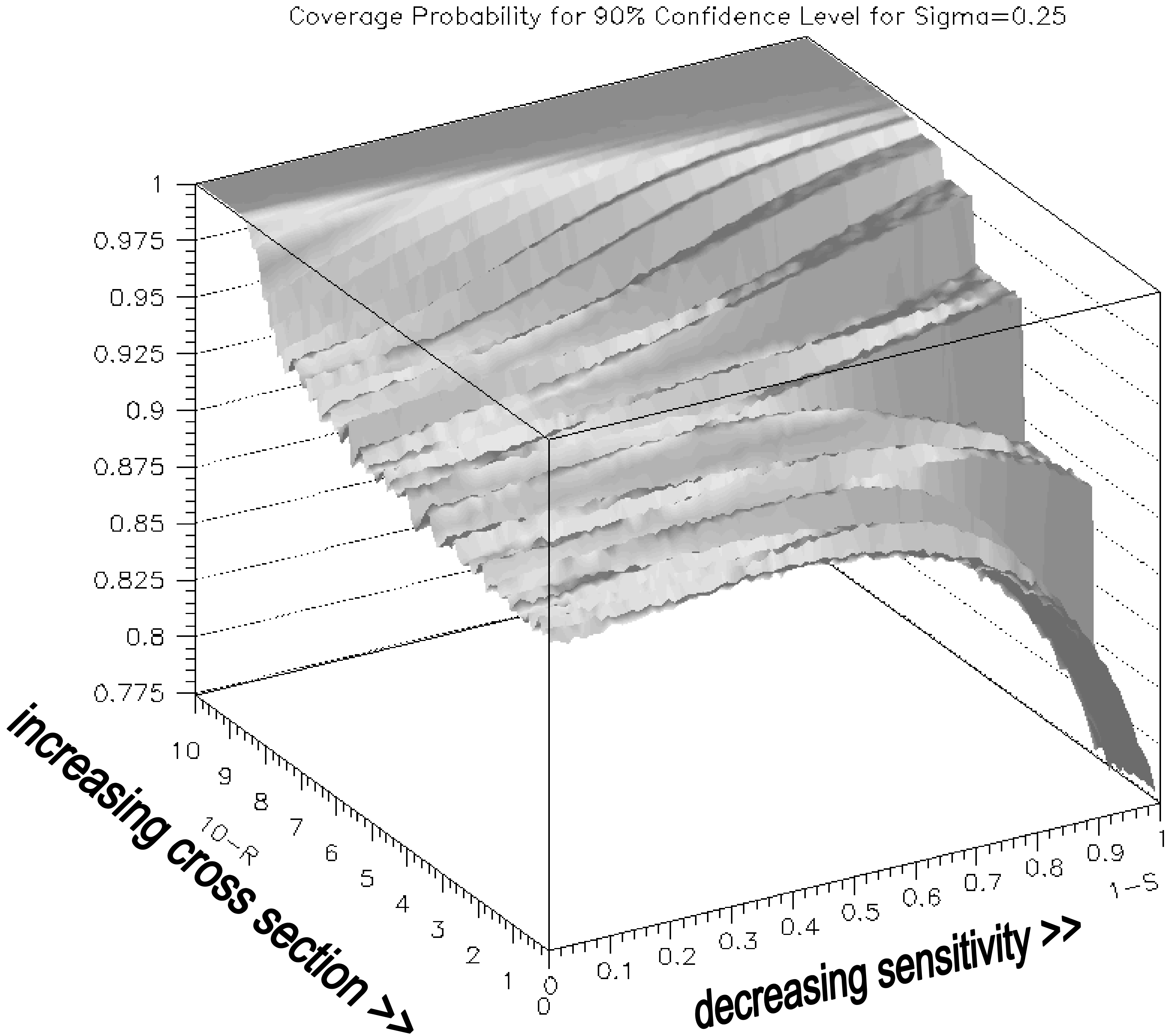
  find % of experiments with limits including σ at each point

- do MC experiments at each value

  pick observed value for sensitivity, k

  calculate limit based on these

  see if limit covers true value of σ



Coverage Probability for 90% Confidence Level for Sigma=0.25
1
0.975
0.95
0.925
0.9
0.875
0.85
0.825
0.8
0.775
10
9
8
7
6
5
4
3
2
1
0
10-R
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1-S
increasing cross section >>
decreasing sensitivity >>

Results for C+H Coverage

- **Fails to cover** for large cross section and small efficiency.

Not too surprising

- a count limit $\mathbf{s^U}$ could be due to any value of σ since $\mathbf{s^U} = \varepsilon L \sigma$
- if sensitivity small, would need a huge $\sigma^{\mathbf{U}}$
- Remember, limit on σ must be valid for

any sensitivity--no matter how improbable

coverage handles statistical fluctuations only

Results for CH Coverage

Note added Jan 2003

- ***I no longer believe the results I presented on CH Coverage***.
- I hadn't understood the plot I chose to show (I mis-interpreted the meaning of the resolution parameter). This particular graph was work of colleagues, though the rest of the results in the talk were mine.
- After discussions with Bob Cousins and Harrison Prosper, we further concluded that the **coverage calculation in our colleagues' internal collaboration note did not implement the CH prescription accurately**.
- It is my present opinion that the coverage of the CH prescription remains an open question.

U = Bayes 95% Upper Limits Credible Interval

- k = number of events observed
- b = expected background
- Defined by integral on posterior probability
- Depends on prior probability for signal

  how to express that we don't know if it exists,

  but would be willing to believe it does?

This is the Faustian part of the bargain!

Posterior: compromise likelihood with prior

Expected coverage of Bayesian intervals

- Theorem: <coverage> = 95% for Bayes 95% interval

  < > = average over (possible) true values weighted by prior

- Frequentist definition is minimum coverage for any value of parameter (especially the true one!)

  not average coverage

- Classic tech support: precise, plausible, misleading

  if true for Poisson, why systematically under cover?

  Because k small is infinitely small part of $[0,\infty]$

  but works beautifully for binomial (finite range)

  - coverage varies with parameter but average is right on
    - "obvious" if you do it with flat prior in parameter

The sadness of Fred James:

JIM, HAVE YOU GONE ASTRAY?

- I am indeed seen to worship at
  Reverend Bayes' establishment
- I'm not a fully baptized member
  - sorry Harrison, not that you haven't tried!
- A skeptical inquirer...or a reluctant convert?

Attraction of treating systematics is great

Is accepting a Prior (*he's uninformative!*) too high a price?

A solution for the tepid?

Can we substitute ***convention*** for ***conviction***?

Either one should be **examined for its consequences!**

Candidate Signal Priors

- **Flat** up to maximum M (e.g. σ_{TOT})
  - (our recommendation--but not invariant!)
  - a convention for BR × cross section
- **$1/\sqrt{s}$** (Jeffreys: reparameterization invariant)

  relatively popular "default" prior
- **1/s** (one of Jeffreys' recommendations)

  get expected posterior mean

  limit invariant under power transformation
- **e^{-as}** not singular at s=0

  Bayes for combining with k=0 prev expt,

  a = relative sensitivity to this experiment

$$P(\sigma|k_0=0,I)=\frac{P(k_0=0|\sigma,I)\times P(\sigma|I)}{\int d\sigma P(k_0=0|\sigma,I)\times P(\sigma|I)}=\frac{\frac{e^{-s_0}}{M}}{\int d\sigma\,\frac{e^{-s_0}}{M}} \tag{A1}$$

where $s_0=\sigma\epsilon_0\mathcal{L}_0$ and we have used $\frac{(s_0+b_0)^0}{0!}=1$. Cancelling constants, and changing the integration variable to s_0 , we find

$$P(\sigma|k=0,I)=\mathcal{L}_0e^{-s_0} \tag{A2}$$

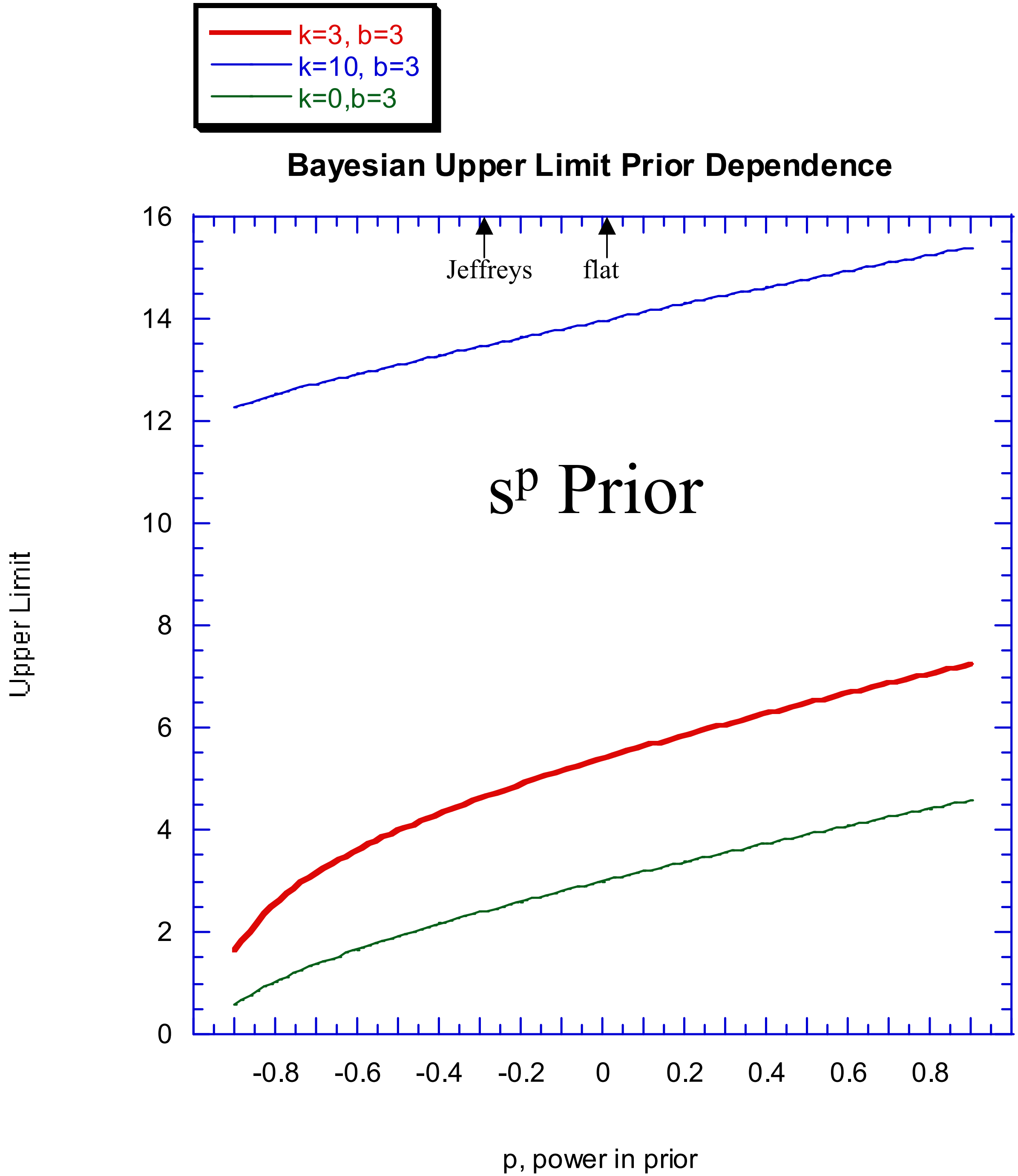
Now consider combining this experiment with a subsequent experiment, with different, but again perfectly known, efficiency, luminosity, and background $\epsilon,\mathcal{L},b$. The natural Bayesian method is to use the posterior for σ from the first experiment as the prior for the second experiment. For the second experiment we write the posterior probability for σ, with k observed events as

$$P(\sigma|k,I)\propto P(k,\sigma,I)\times P(\sigma|I)=e^{-s}\frac{(s+b)^k}{k!}\mathcal{L}_0e^{-s_0} \tag{A3}$$

using $s=\sigma\epsilon\mathcal{L}$. Now we write s_0 in terms of s by recognizing

$$s_0=\sigma\epsilon_0\mathcal{L}_0=\sigma\epsilon\mathcal{L}\frac{\epsilon_0\mathcal{L}_0}{\epsilon\mathcal{L}}=s\frac{\epsilon_0\mathcal{L}_0}{\epsilon\mathcal{L}}=as \tag{A4}$$

$$P(\sigma|k,I)\propto e^{-s}\frac{(s+b)^k}{k!}e^{-as}=e^{-s}\frac{(s+b)^k}{k!}e^{-\epsilon_0\mathcal{L}_0\sigma} \tag{A6}$$


k=3, b=3
k=10, b=3
k=0,b=3
Bayesian Upper Limit Prior Dependence
Jeffreys
flat
s^p Prior
Upper Limit
p, power in prior
16
14
12
10
8
6
4
2
0
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8

Power Family s^p Results (δ_b=0)

- The **flat prior is not "special"** (stationary)

  But if b=0, Bayes UL = Frequentist UL → coverage

  but lower limit would differ

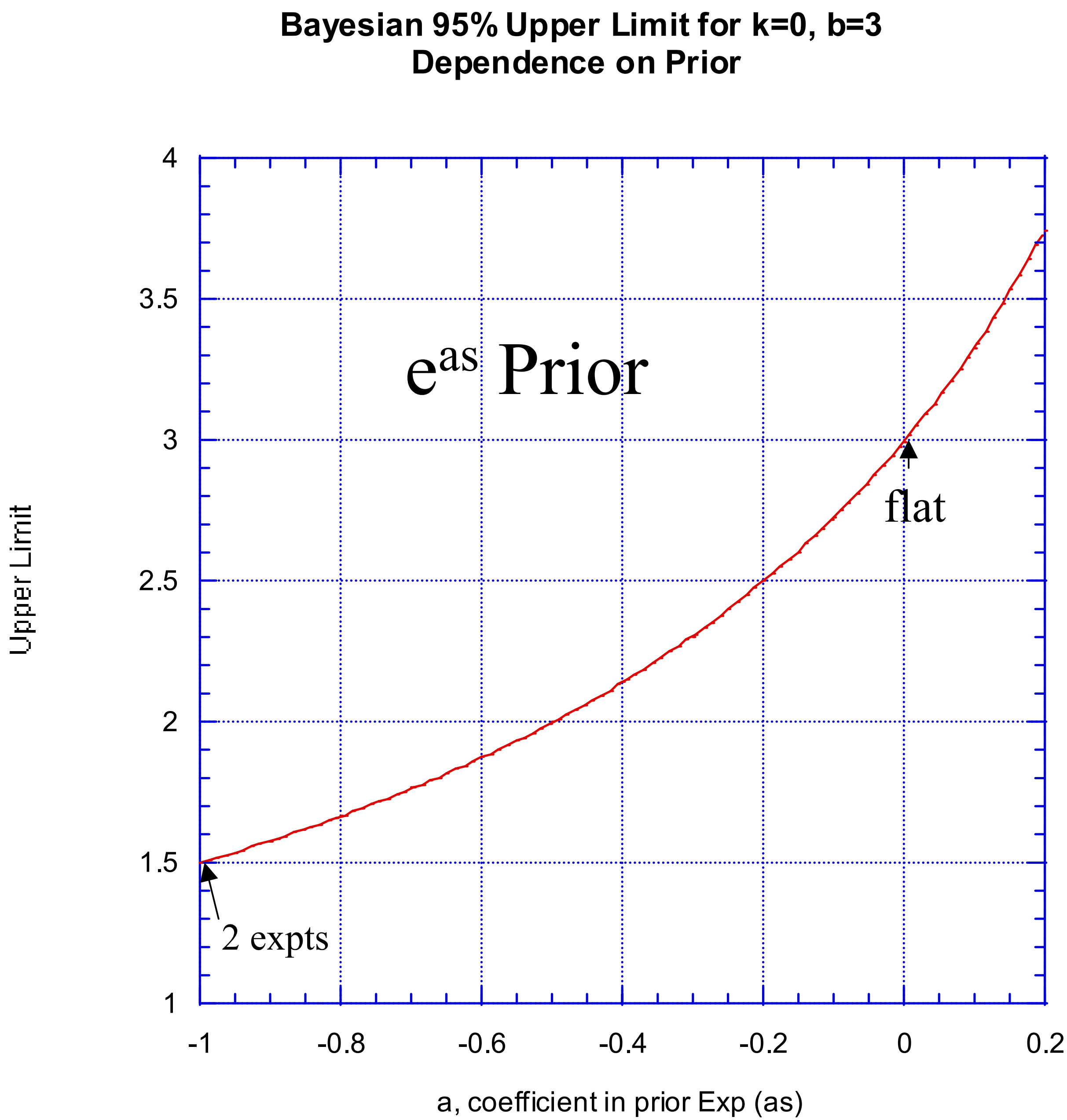
- **$1/\sqrt{s}$** gives **smaller limit (more weight to s=0)**
  – less coverage than flat (though converges for $k \to \infty$)
- **1/s** gives you **0** upper limit if b > 0

  too prejudiced towards 0 signal!

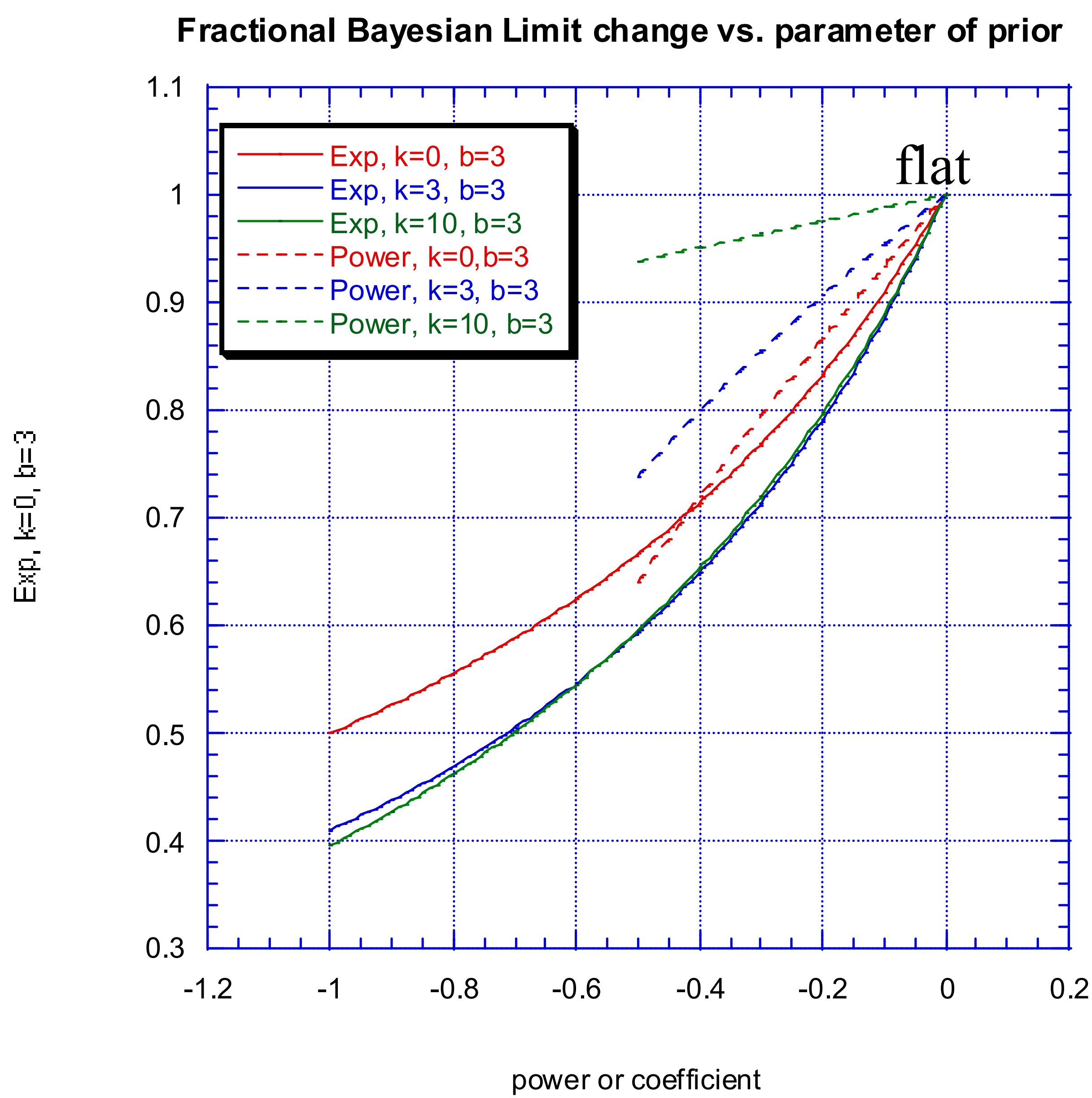
- More p dependence for k=0 than k=3 or k=10

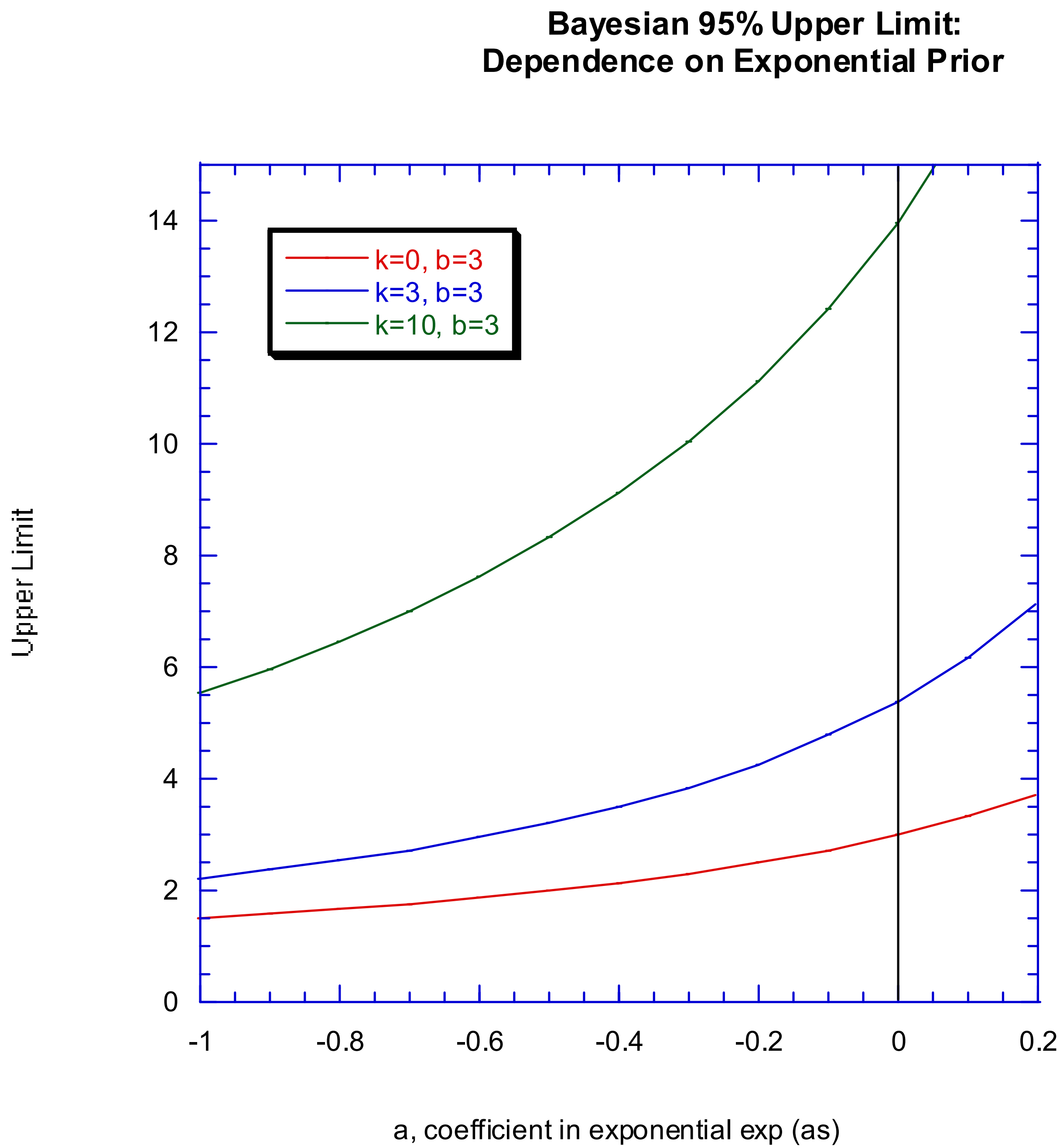
  flat (p=0) to $1/\sqrt{s}$ gives 36%, 26% , 6%

  data able to overwhelm prior (b=3)



Bayesian 95% Upper Limit for k=0, b=3
Dependence on Prior
e^as Prior
flat
2 expts
Upper Limit
a, coefficient in prior Exp (as)
4
3.5
3
2.5
2
1.5
1
-1
-0.8
-0.6
-0.4
-0.2
0
0.2


Fractional Bayesian Limit change vs. parameter of prior
Exp, k=0, b=3
Exp, k=3, b=3
Exp, k=10, b=3
Power, k=0,b=3
Power, k=3, b=3
Power, k=10, b=3
flat
Exp, k=0, b=3
power or coefficient
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
-1.2
-1
-0.8
-0.6
-0.4
-0.2
0
0.2



Bayesian 95% Upper Limit:
Dependence on Exponential Prior
k=0, b=3
k=3, b=3
k=10, b=3
Upper Limit
a, coefficient in exponential exp (as)
0
2
4
6
8
10
12
14
-1
-0.8
-0.6
-0.4
-0.2
0
0.2

Exponential Family Results (δ_b=0)

- Peak at s=0 pulls limit lower than flat prior
- effects larger than $1/\sqrt{s}$ vs. flat: equivalent to *data*
- e^{-s} gives you 1/2 the limit of flat (a=0) for k=0:

  combined 2 equal experiments

- biggest fractional effects on k=10   (=1/2.5)

  because disagrees with previous k=0 measurement

  opposite tendency of power family

  k=10 least dependent on power

Dependence on Efficiency **Informative Prior**
(representation of systematics)

- Input: **estimated efficiency** and **uncertainty**

  $\eta \equiv$ uncertainty/estimate

  "efficiency" is really εL (a **nuisance** parameter)

- Consider forms for efficiency prior

  Expect: less fractional dependence on form of prior

  - than on signal prior form
  - because of the constraint of the input: informative

- study using flat prior for cross section, $\delta_b=0$
- Warning: $s = \varepsilon L \times \sigma$ (multiplicative form)

  limit in s could mean low efficiency **or** high σ

Expressing $\hat{\varepsilon} \pm \delta\varepsilon$

$\eta \equiv \delta\varepsilon / \hat{\varepsilon}$

- "obvious" Truncated Gaussian (Normal)
  - model for additive errors
  - we recommend(ed)
  - truncate so efficiency ≥ 0
- **Lognormal** (Gaussian in Ln ε)
  - model for multiplicative errors
- **Gamma** (Bayes conjugate prior)
  - flat prior + estimate of Poisson variable
- **Beta** (Bayes Conjugate prior)
  - flat prior + estimate of Binomial variable

$$\eta = \sqrt{\left(\frac{\delta_\epsilon}{\hat{\epsilon}}\right)^2 + \left(\frac{\delta_{\mathcal{L}}}{\hat{\mathcal{L}}}\right)^2}. \tag{6.8}$$

For the purposes of this section, it is convenient to define a scaled sensitivity variable

$$\phi = \epsilon\mathcal{L}/\hat{\epsilon}\hat{\mathcal{L}} \tag{6.9}$$

where $\hat{\phi} = 1 \pm \eta$. In this spirit, we will use η to parameterize the informative prior for ϕ, rather than adjusting the posterior mean and rms of this distribution to precisely match the estimates. Without loss of generality, we can further consider unit expected sensitivity $\hat{\epsilon}\hat{\mathcal{L}} = 1$, so that $s = \mathcal{L}\epsilon\sigma = \hat{\mathcal{L}}\hat{\epsilon}\phi\sigma = \phi\sigma$ and we can easily compare numerical values of the upper limits with other results. In the usual fashion, the posterior probability for the cross section will be given by

$$P(\sigma|k) \propto P(\sigma) \int d\phi P(k|\phi\sigma + b) P(\phi|\eta) \tag{6.10}$$

$$TGauss(\phi|\eta) = \frac{1}{2\pi\eta} \exp -\frac{1}{2}\left(\frac{\phi - 1}{\eta}\right)^2 \tag{6.11}$$

$$lNor(\phi|\eta) = \frac{1}{\phi 2\pi\eta} \exp -\frac{1}{2}(\ln \phi/\eta)^2 \tag{6.12}$$

$$Gamma(\phi|\eta) \propto \phi^{1/\eta^2} e^{-\phi/\eta^2} \tag{6.13}$$

$$Beta(\epsilon; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \epsilon^{a-1}(1-\epsilon)^{b-1}. \quad (6.14)$$

The estimate efficiency and uncertainty are assumed to have come from $\hat{\epsilon} = K/N$, the fraction of successes, and $\delta_{\epsilon} = \eta\hat{\epsilon} = \sqrt{\hat{\epsilon}(1-\hat{\epsilon})/N}$. From these, the parameters can be deduced by

$$N = \hat{\epsilon}(1-\hat{\epsilon})/\delta_{\epsilon}^2 = (1-\hat{\epsilon})/(\eta^2\hat{\epsilon}) \quad (6.15)$$

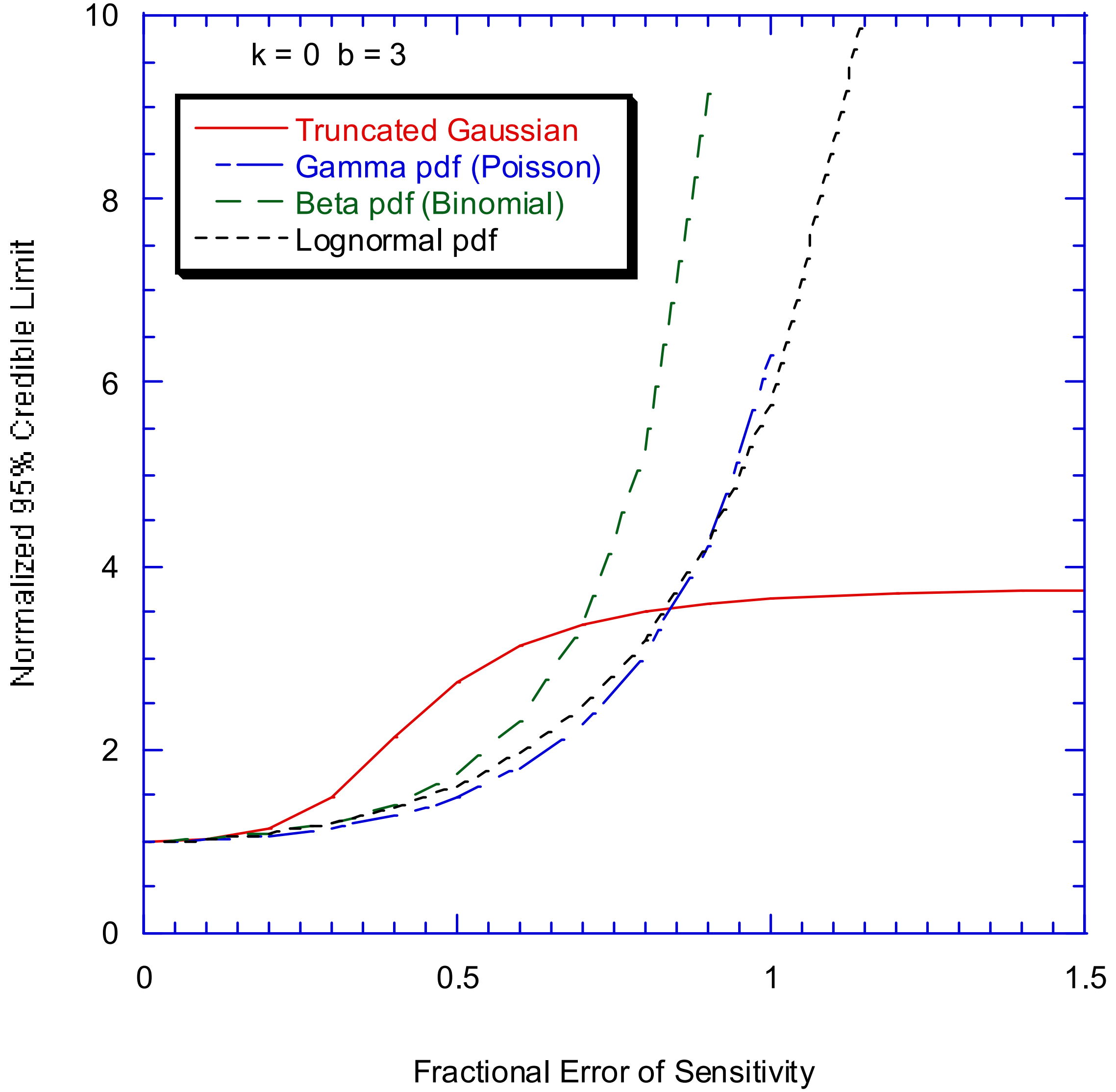
and (note the convergence to the Poisson case for $\hat{\epsilon} \to 0$)

$$K = \hat{\epsilon}N = (1-\hat{\epsilon})/\eta^2 \quad (6.16)$$

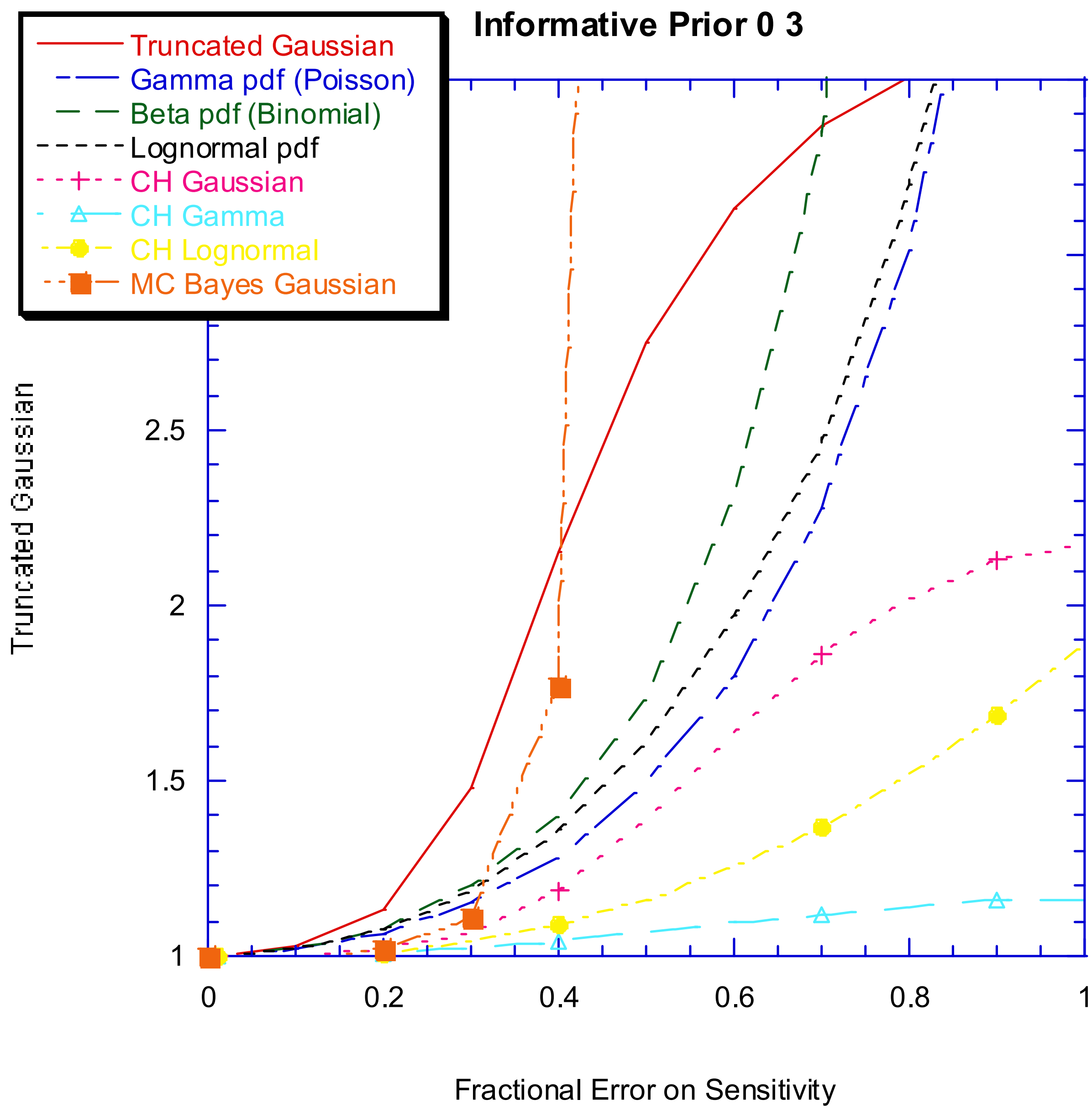
resulting in

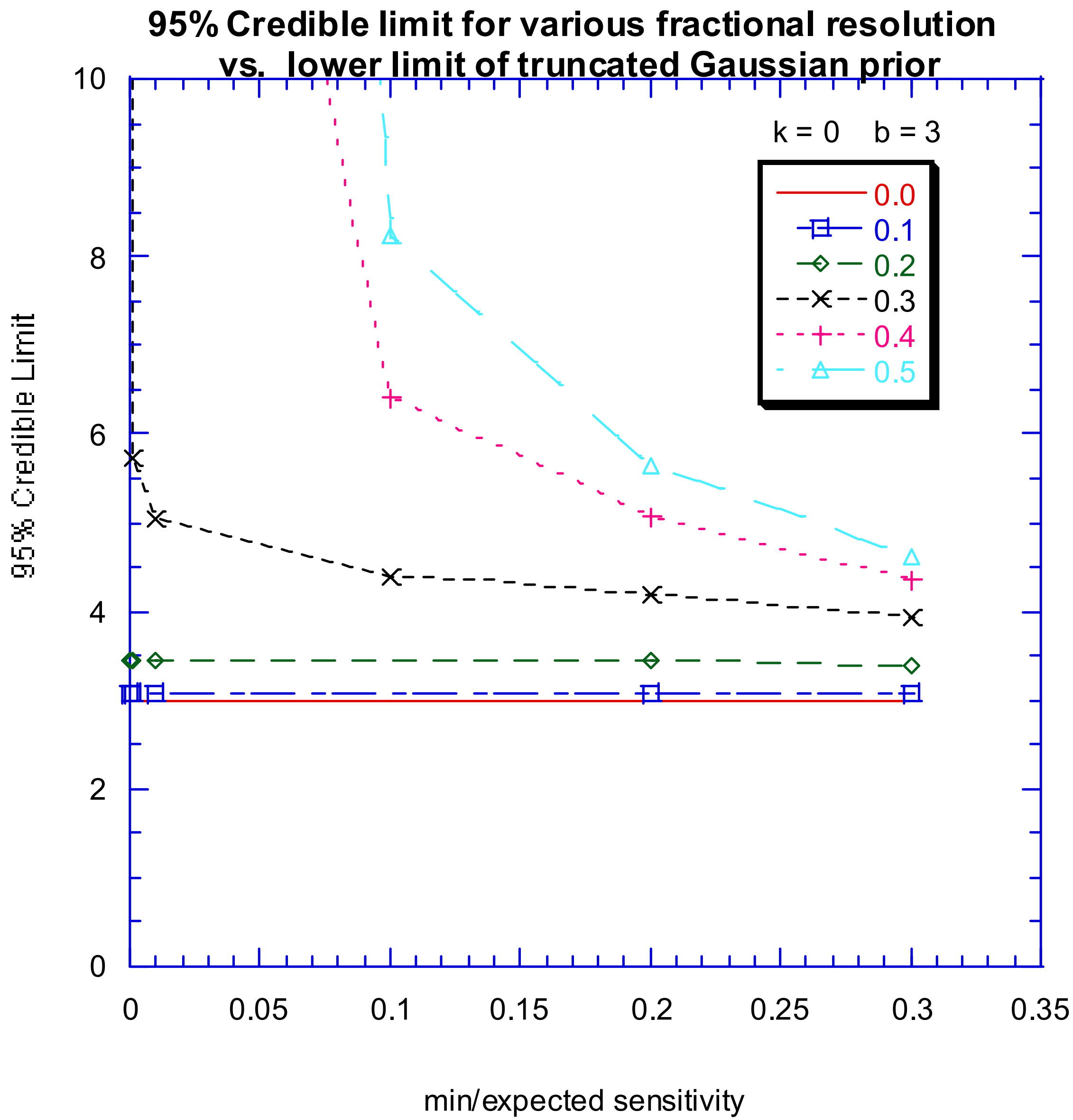
$$a = 1 + K = (1-\hat{\epsilon})/\eta^2, \quad (6.17)$$

$$b = 1 + N - K = 1 + (1/\hat{\epsilon} - 1)/(1-\hat{\epsilon})/\eta^2 \quad (6.18)$$



k = 0 b = 3
Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
Normalized 95% Credible Limit
Fractional Error of Sensitivity
0
2
4
6
8
10
0
0.5
1
1.5


Informative Prior 0 3
Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
CH Gaussian
CH Gamma
CH Lognormal
MC Bayes Gaussian
Truncated Gaussian
2.5
2
1.5
1
0
0.2
0.4
0.6
0.8
1
Fractional Error on Sensitivity



95% Credible limit for various fractional resolution
vs. lower limit of truncated Gaussian prior
k = 0 b = 3
0.0
0.1
0.2
0.3
0.4
0.5
95% Credible Limit
min/expected sensitivity
0
2
4
6
8
10
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35

Results for Truncated Gaussian

- A bad choice, especially if $\eta > .2$ or so
- cutoff-dependent (MC: 4 sigma; calc .1$<\varepsilon>$)

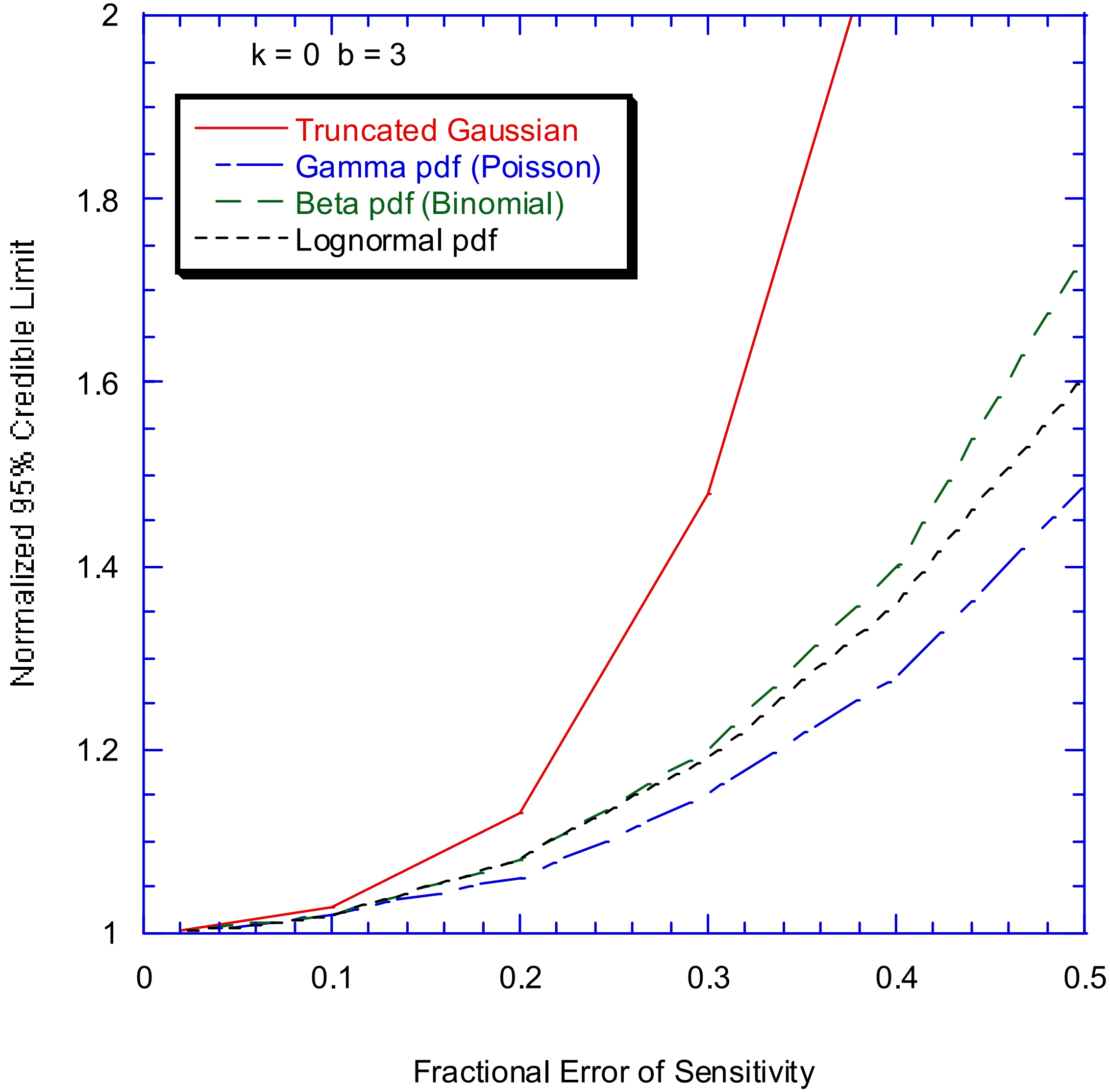
  Otherwise depends on M, range of prior for σ
- MC of course cranks out some answer
  - dependent on luck, and cutoffs of generators
- WHY!? (same problem as with Coverage)
  - **Can't set limit if possibility of no sensitivity**

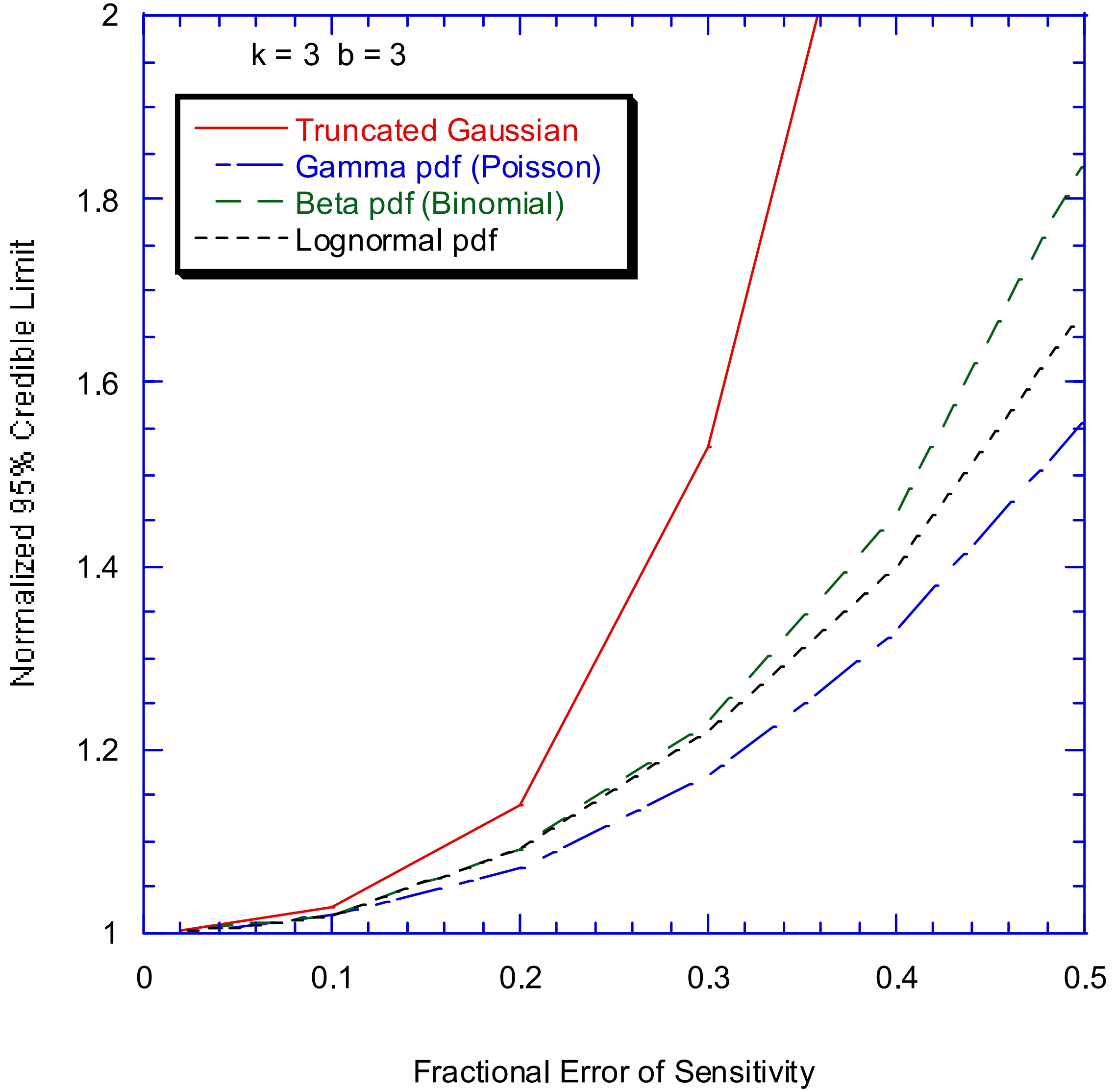
  Probability of $\varepsilon=0$ always finite for a truncated Gaussian

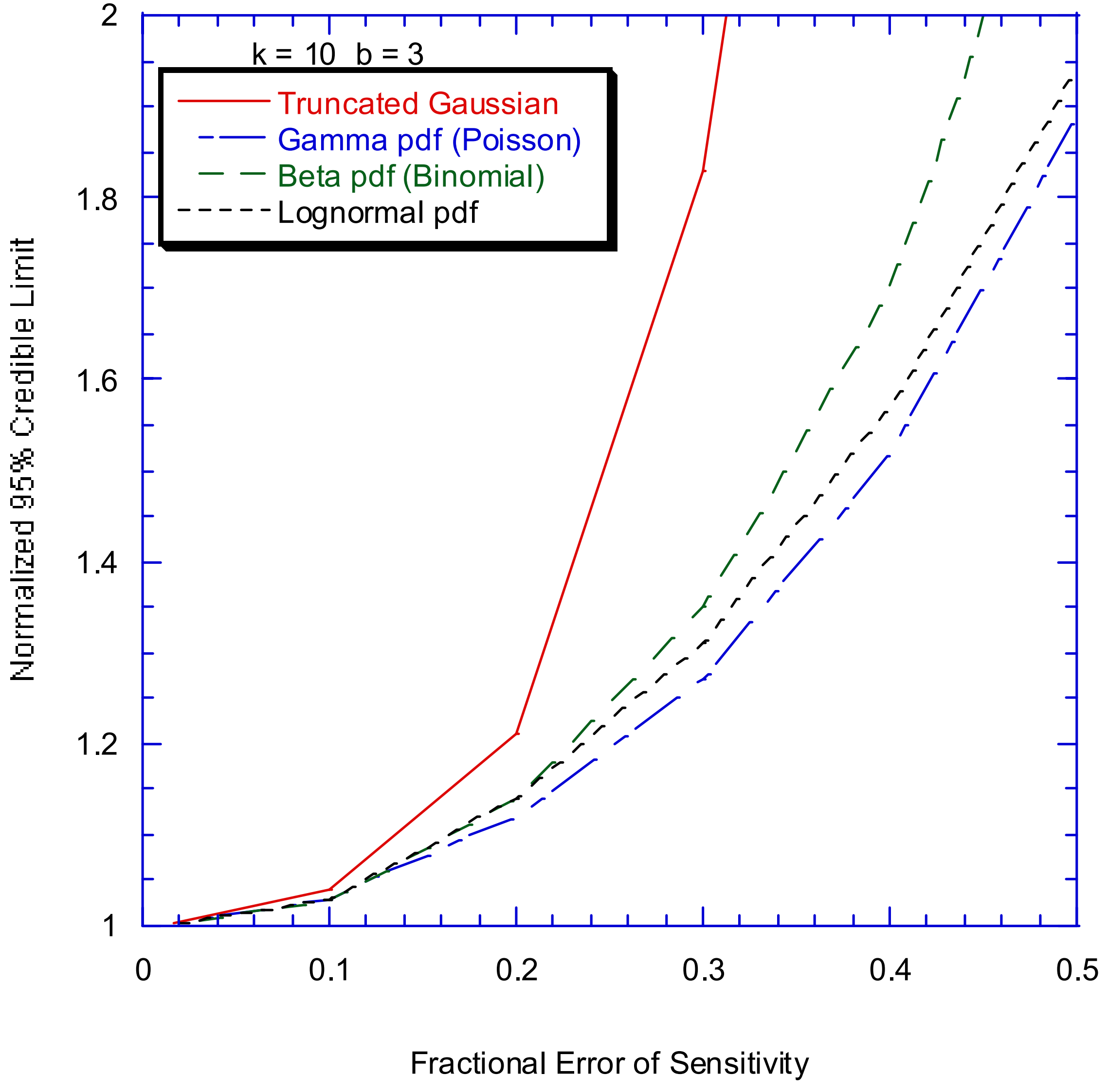
  with flat prior in σ, gives long tail in σ posterior

  Bayes takes this literally:

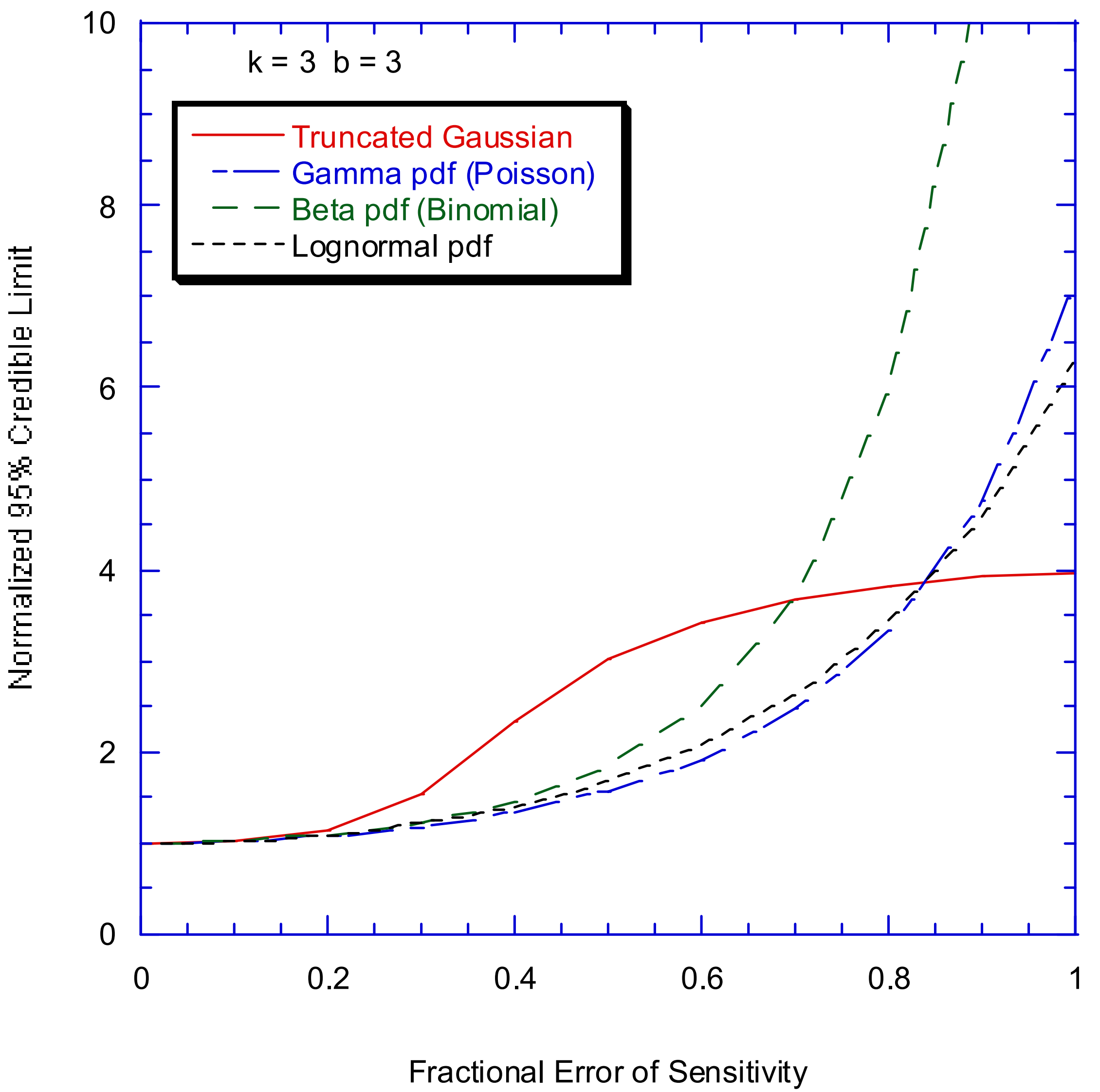
  U reflects heavy weighting of large cross section!


k = 0 b = 3
Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
Normalized 95% Credible Limit
Fractional Error of Sensitivity
2
1.8
1.6
1.4
1.2
1
0
0.1
0.2
0.3
0.4
0.5


k = 3 b = 3
Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
Normalized 95% Credible Limit
Fractional Error of Sensitivity
2
1.8
1.6
1.4
1.2
1
0
0.1
0.2
0.3
0.4
0.5



k = 10 b = 3
Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
Normalized 95% Credible Limit
Fractional Error of Sensitivity
2
1.8
1.6
1.4
1.2
1
0
0.1
0.2
0.3
0.4
0.5



k = 3 b = 3
Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
Normalized 95% Credible Limit
Fractional Error of Sensitivity
0
2
4
6
8
10
0
0.2
0.4
0.6
0.8
1

Results for alternatives

ALL have $P(\varepsilon=0) = 0$ naturally

- Lognormal, beta, and gamma

  not very different (as expected--informative)

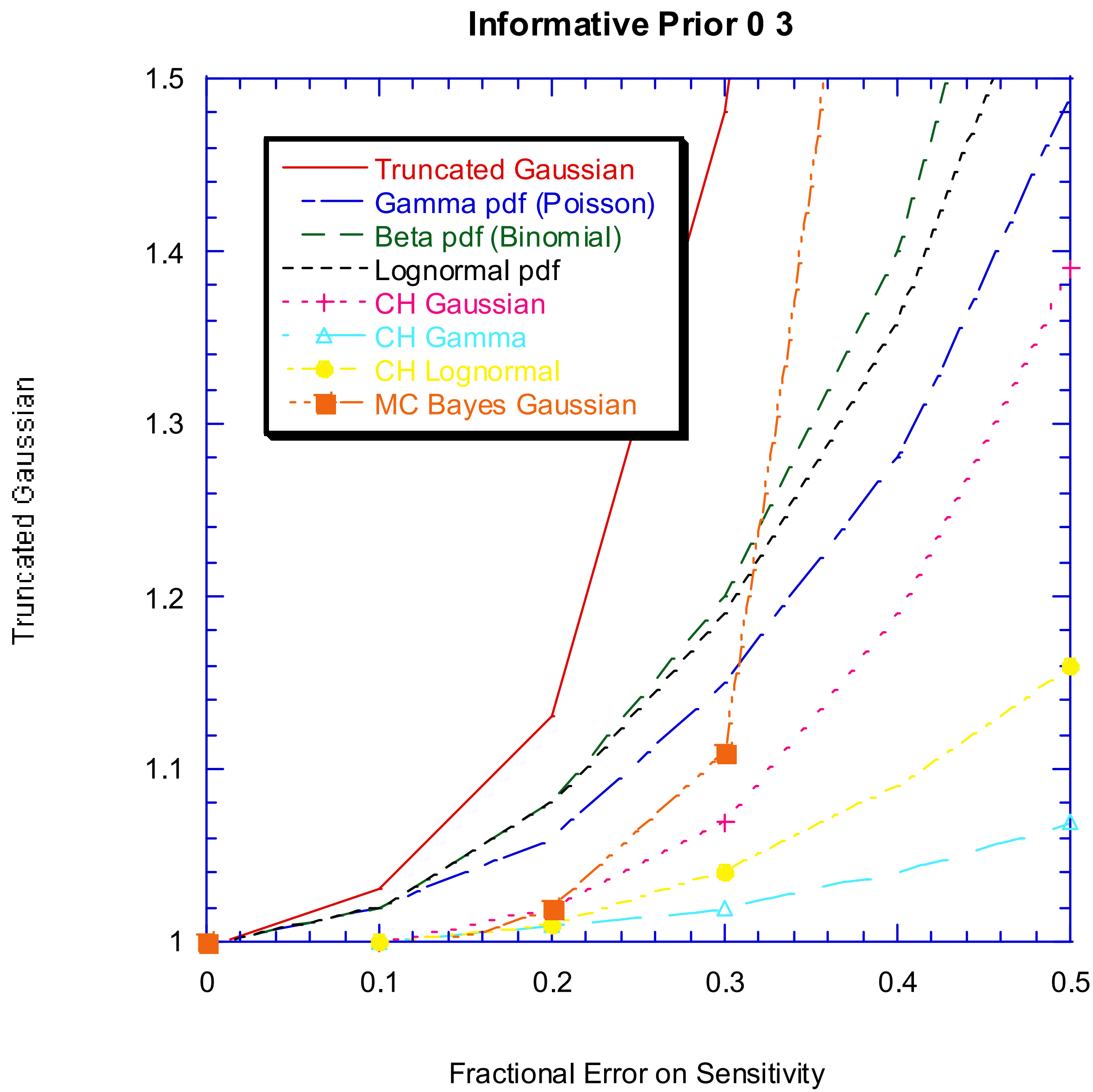
  opinion: comparable to "choice of ensemble"

- Not a Huge effect:

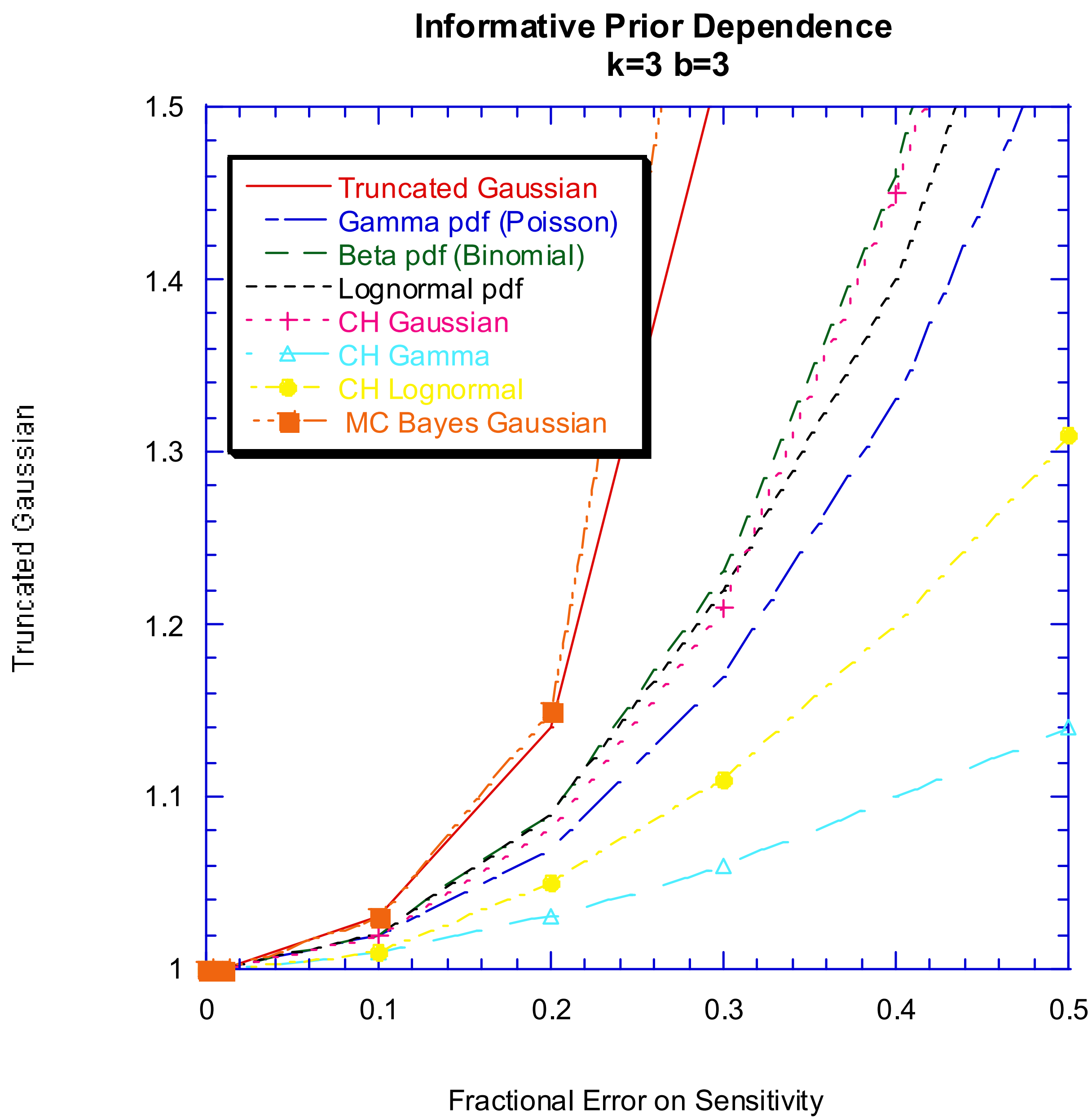
  $U(\eta)/U(0) < 1+\eta$ up to $\eta \sim 1/3$

  . . .

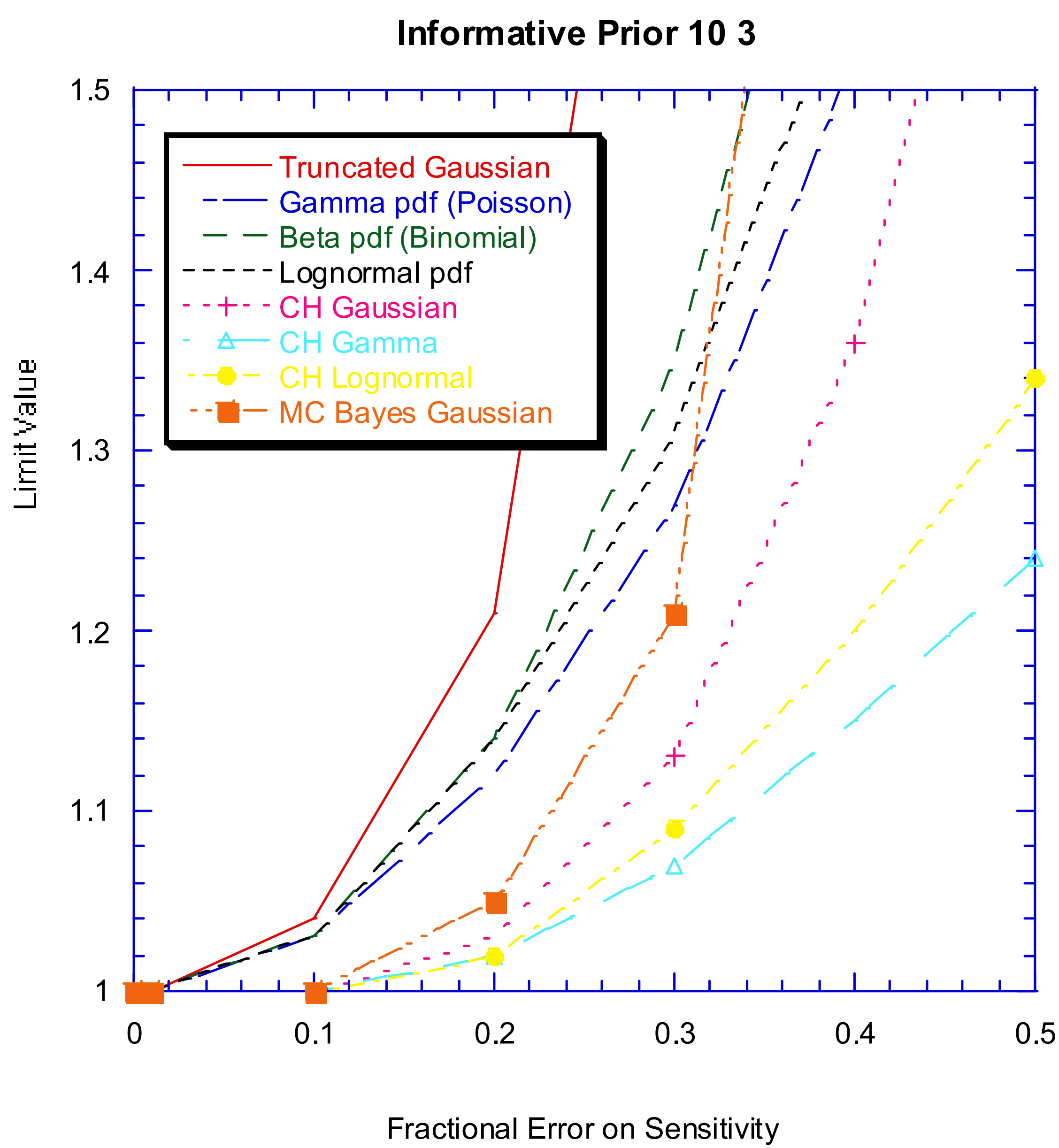
- Lognormal, Gamma can be expressed as efficiency scaled to 1.0 (so can Gaussian)
- beta requires absolute scale $(1-\varepsilon)^j$



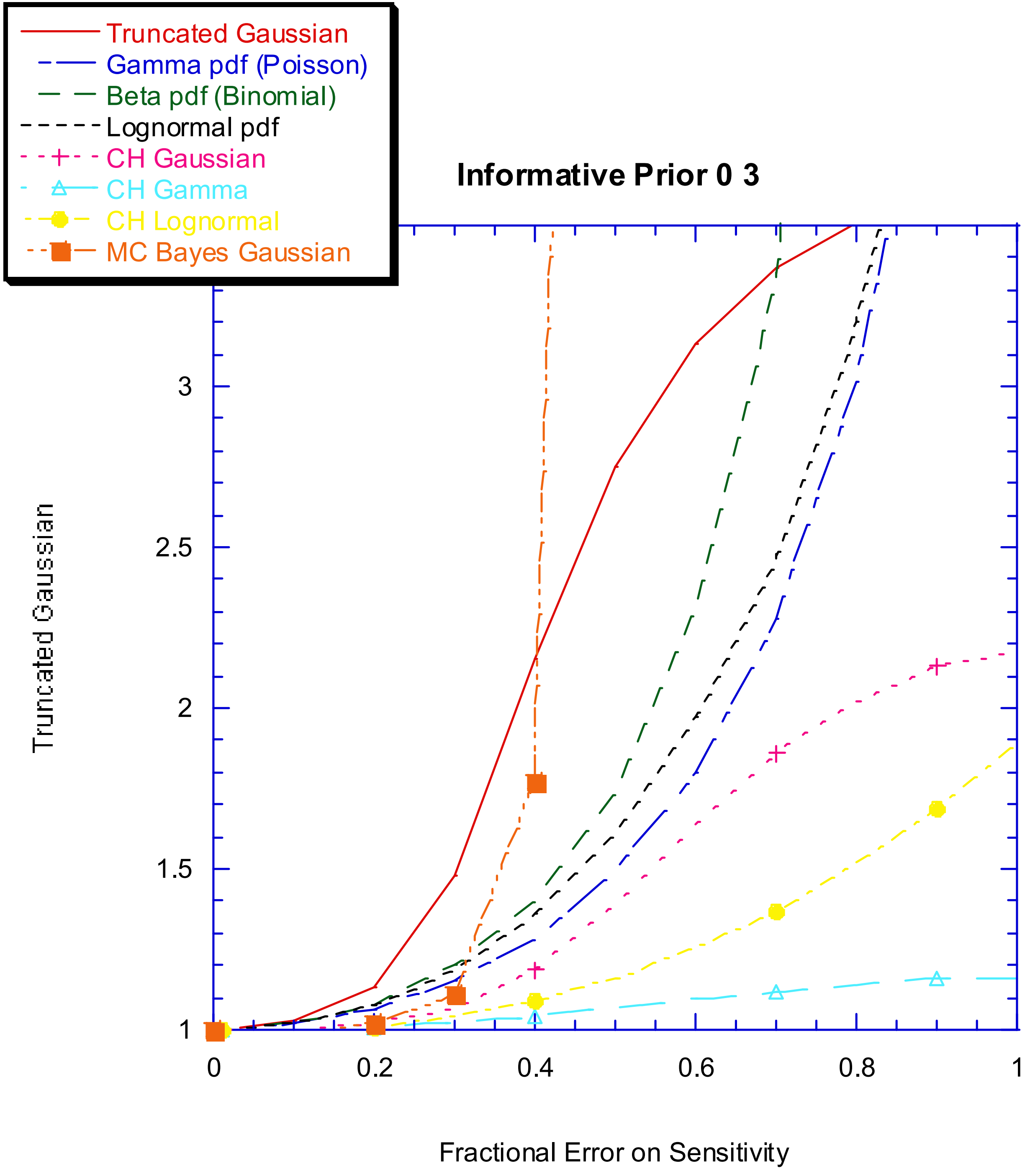
Informative Prior 0 3
Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
CH Gaussian
CH Gamma
CH Lognormal
MC Bayes Gaussian
1.5
1.4
1.3
1.2
1.1
1
0
0.1
0.2
0.3
0.4
0.5
Truncated Gaussian
Fractional Error on Sensitivity

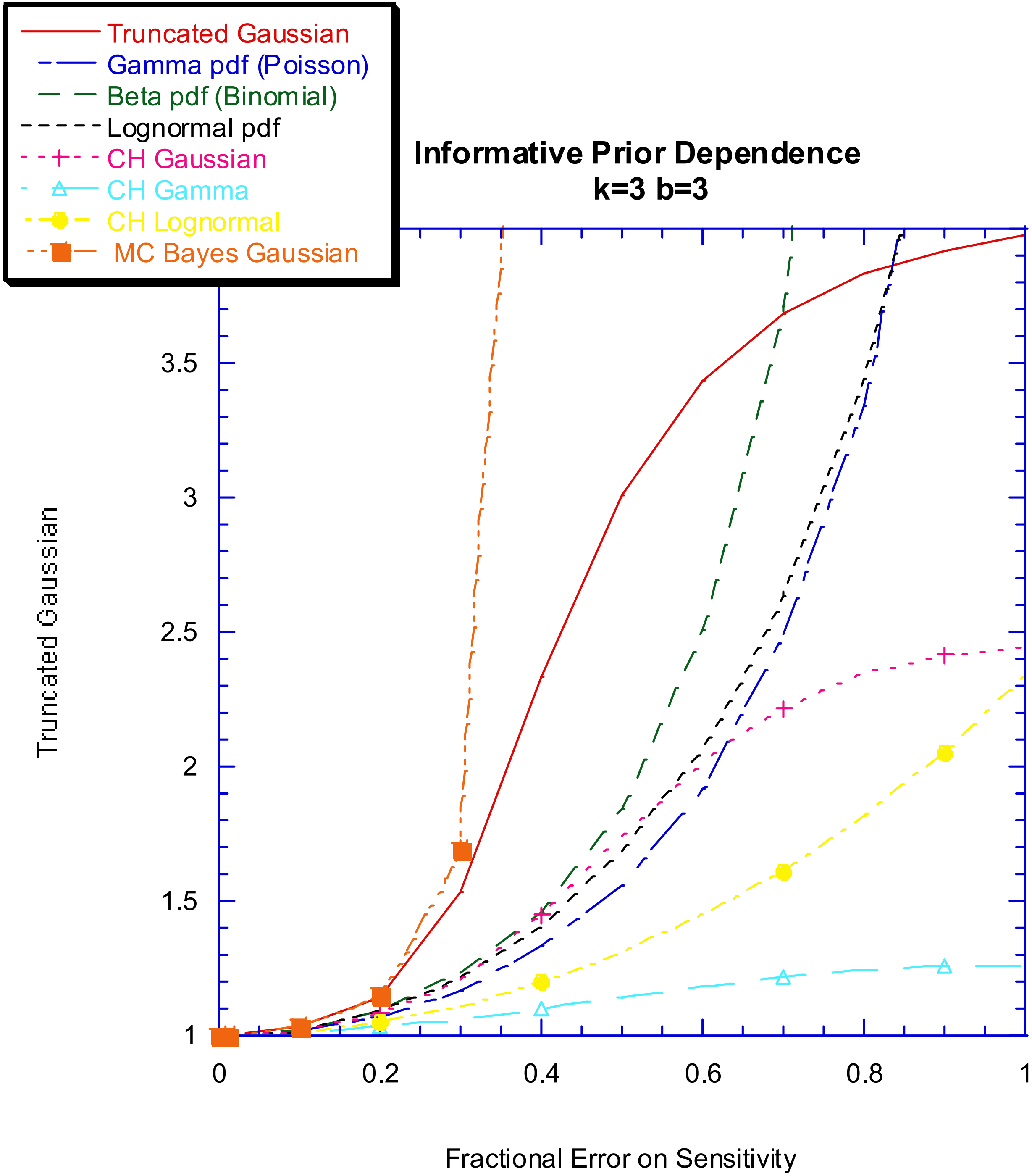


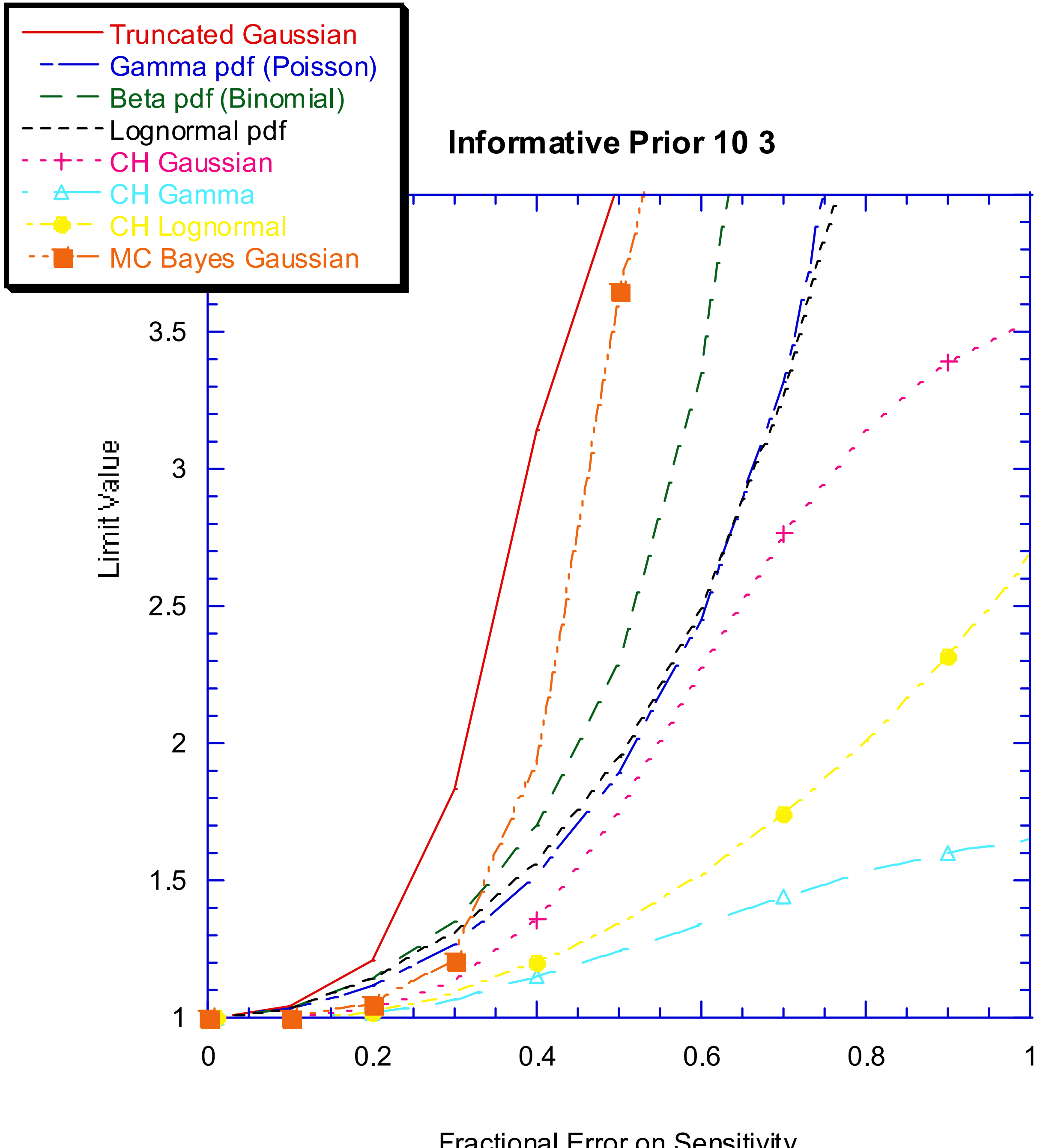
Informative Prior Dependence
k=3 b=3
Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
CH Gaussian
CH Gamma
CH Lognormal
MC Bayes Gaussian
1.5
1.4
1.3
1.2
1.1
1
0
0.1
0.2
0.3
0.4
0.5
Truncated Gaussian
Fractional Error on Sensitivity



Informative Prior 10 3
Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
CH Gaussian
CH Gamma
CH Lognormal
MC Bayes Gaussian
Limit Value
1.5
1.4
1.3
1.2
1.1
1
0
0.1
0.2
0.3
0.4
0.5
Fractional Error on Sensitivity


Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
CH Gaussian
CH Gamma
CH Lognormal
MC Bayes Gaussian
Informative Prior 0 3
Truncated Gaussian
Fractional Error on Sensitivity
1
1.5
2
2.5
3
0
0.2
0.4
0.6
0.8
1


Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
CH Gaussian
CH Gamma
CH Lognormal
MC Bayes Gaussian
Informative Prior Dependence
k=3 b=3
Truncated Gaussian
3.5
3
2.5
2
1.5
1
0
0.2
0.4
0.6
0.8
1
Fractional Error on Sensitivity


Truncated Gaussian
Gamma pdf (Poisson)
Beta pdf (Binomial)
Lognormal pdf
CH Gaussian
CH Gamma
CH Lognormal
MC Bayes Gaussian
Informative Prior 10 3
Limit Value
3.5
3
2.5
2
1.5
1
0
0.2
0.4
0.6
0.8
1
Fractional Error on Sensitivity

Results, compared with C+H (mixed Frequentist-Bayes)

- Truncated Gaussian well-behaved for C+H

  no flat prior to compound with $P(\varepsilon=0) > 0$?

  Fairly close to Bayes Lognormal

- **C+H Limits depend on form** of informative prior MORE than Bayes

  Lognormal, gamma C+H lower than Bayes!

- C+H limits lower than Bayes limits

  Which is "better"? coverage study?

  C+H Gaussian undercovers for small ε ($\rightarrow$large σ)

  But: I now (Jan '03) believe this remains an open question

Dependence on Background Uncertainty

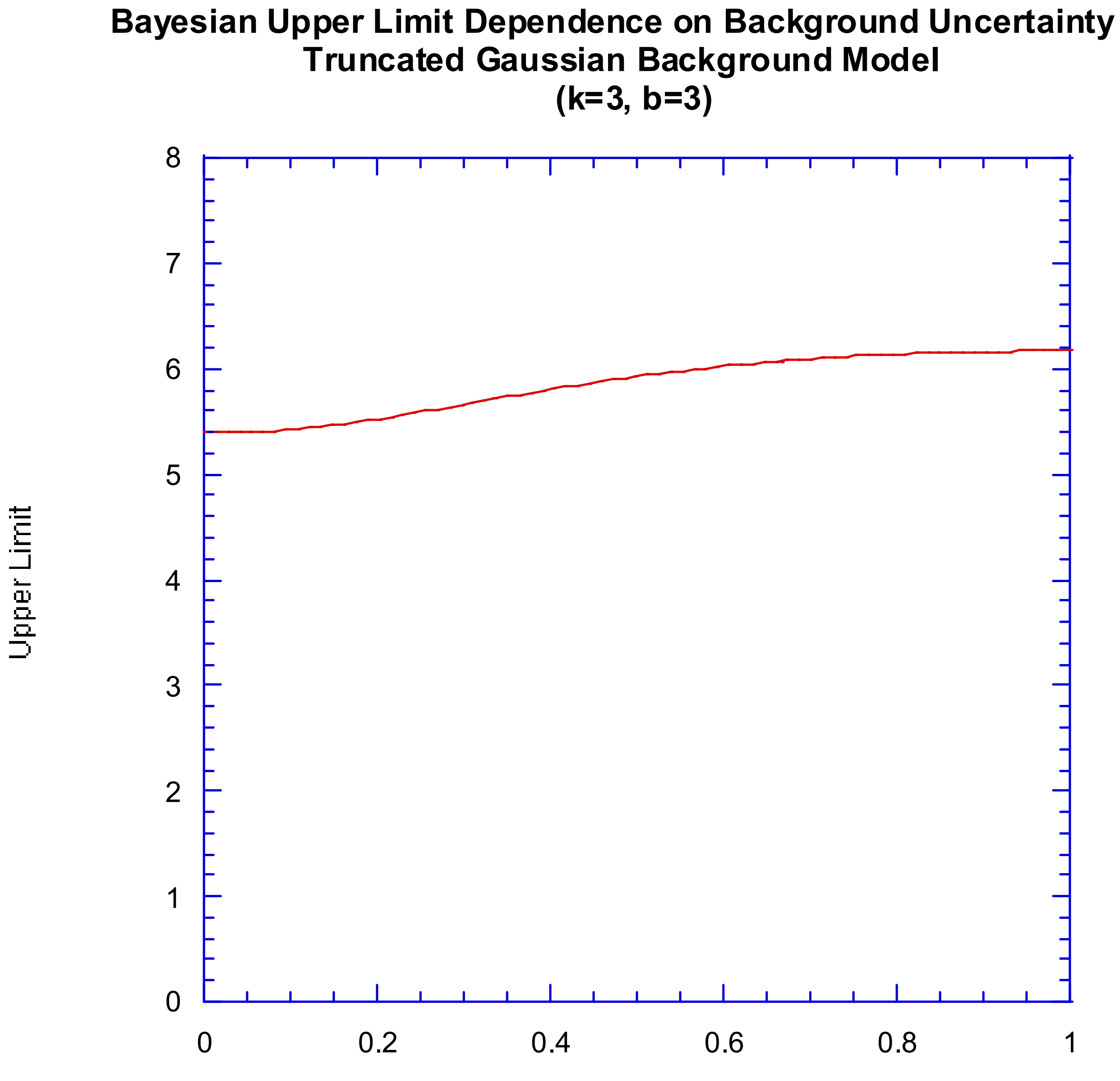
- Use flat prior, no efficiency uncertainty
- Use truncated Gaussian to represent $\langle b \rangle \pm \delta_b$

But isn't that a disaster? No--

additive is very different from multiplicative

$$\varepsilon L \sigma + \mathbf{b}$$

behavior at b=0 not special


Bayesian Upper Limit Dependence on Background Uncertainty
Truncated Gaussian Background Model
(k=3, b=3)
Upper Limit
db/b
0
1
2
3
4
5
6
7
8
0.2
0.4
0.6
0.8
1

Background Prior Results

- Result: **very mild dependence** on $\pm\delta_b/b$

< 10% change up to $\delta_b/b = .66$

most sensitive for k=3, b=3; k=1, b=3

absolute maximum: set b=0   20-40% typically

set b=0: force Frequentist coverage?

- No need to consider more complex models

Paper in preparation

- With Harrison Prosper and Marc Paterno
  coverage calculation: more DØ help

- Thanks to Louis Lyons for the prod to finish
  - and a 2nd chance at understanding all this
    - only 1 hour jet lag, maybe I'll be awake

- Poisson, Fisher….

Summary

(out of things to say)

Cases studied: b=3, k=0,3,10 mostly

studies changed one thing at a time

- ***All Bayes upper limits seen to monotonically increase with uncertainties***

(couldn't quite prove:

Goedel's Theorem for Dummies)

Hello PDG/RPP

nuisance effects 15% or so--please advise us

ignoring them gives too-optimistic limits

Signal Prior Summary

Flat signal prior a convention

b=0, η=0 matches Frequentist upper limit

we still recommend it

careful it's not normalized

flat vs 1/√s **matters at 30% level** when setting limits

So publish what you did!

Enough info to deduce **$N^U = \sigma^U/\langle\varepsilon L\rangle$** at one point

can see if method or results differ

how about posting limits programs on web?

exponential family actually is a strong opinion (=data)

Informative Prior Summary

Can't set limit if possibility of no sensitivity

- C+H mixed prescription doesn't cover
    - Note added Jan 2003: I now believe this remains an open question!
  - how well does Bayes do? ("better"?)
- Efficiency informative prior **matters** in Bayesian

  at a **level of 10%** differences if you avoid Gaussian

  Prefer Lognormal over Truncated Gaussian

  Keep uncertainty under 30% (large, ill-defined!)
    - limit grows 20-30% for 30% fractional error in efficiency
    - growth worse than quadratic

  Bayesian upper limits larger than C+H; more similar

  Publish what you did
- Background uncertainty weaker effect than efficiency
  - typically < 15% even at $\delta b/b=1$

Is 20% difference in limits worth a religious war ...?

(less of a problem if we actually find something!)

- Flat σ Prior broadly useful in counting expts?
- Set limits on visible cross section $\sigma^U(\theta)$

  signal MC for ε (θ)

  stays as close as we can get to raw counts

  here is where scheme-dependence hits; it's not too bad…

  resolution corrections, prior dependence ~ 20-30% or less

- Interpret exclusion limits for θ:

  compare σ^U to $\sigma(\theta)$

  IF steep parameter dependence: less scheme-dependence

  in limits for θ than $\sigma^U(\theta)$...