Statistical Issues in Particle Physics

Intensive Research Session at SAMSI March 6–10 2006

the Bayesian approach to setting limits: what to avoid

Joel Heinrich

University of Pennsylvania

Collider Detector at Fermilab (CDF) Collaboration

The task of setting limits in situations involving nuisance parameters with uncertainties has proved a difficult one in practice. CDF's Statistics Committee has recently recommended a Bayesian approach to setting limits.

While investigating the performance of that approach, one rather restricted scenario was found to result in poor coverage behavior. The scenario is described, the resulting poor coverage behavior is illustrated, and solutions are proposed.

1st test case: upper limit for single channel Poisson process

Observe n events from a process with Poisson rate $\boxed{\epsilon s + b}$, where s is cross section, ϵ is acceptance$\times$luminosity, b is background, and obtain the Bayesian posterior for s. Nuisance parameters ϵ and b are determined via Poisson subsidiary measurements, whose posteriors serve as the priors for ϵ and b in the main measurement. The specified Bayesian priors are

- flat prior for $s > 0$
- flat (subsidiary) prior for $\epsilon > 0$
- flat (subsidiary) prior for $b > 0$

After marginalizing over ϵ and b, we obtain an upper limit for s by integrating the posterior with respect to s from $s = 0$ to the value of s that yields credibility level β.

the subsidiary measurements

The subsidiary measurement for ϵ observes m events with Poisson rate $\kappa\epsilon$, where κ is a known constant. The subsidiary posterior,

$$p(\epsilon|m) = \frac{\kappa(\kappa\epsilon)^m e^{-\kappa\epsilon}}{m!}$$

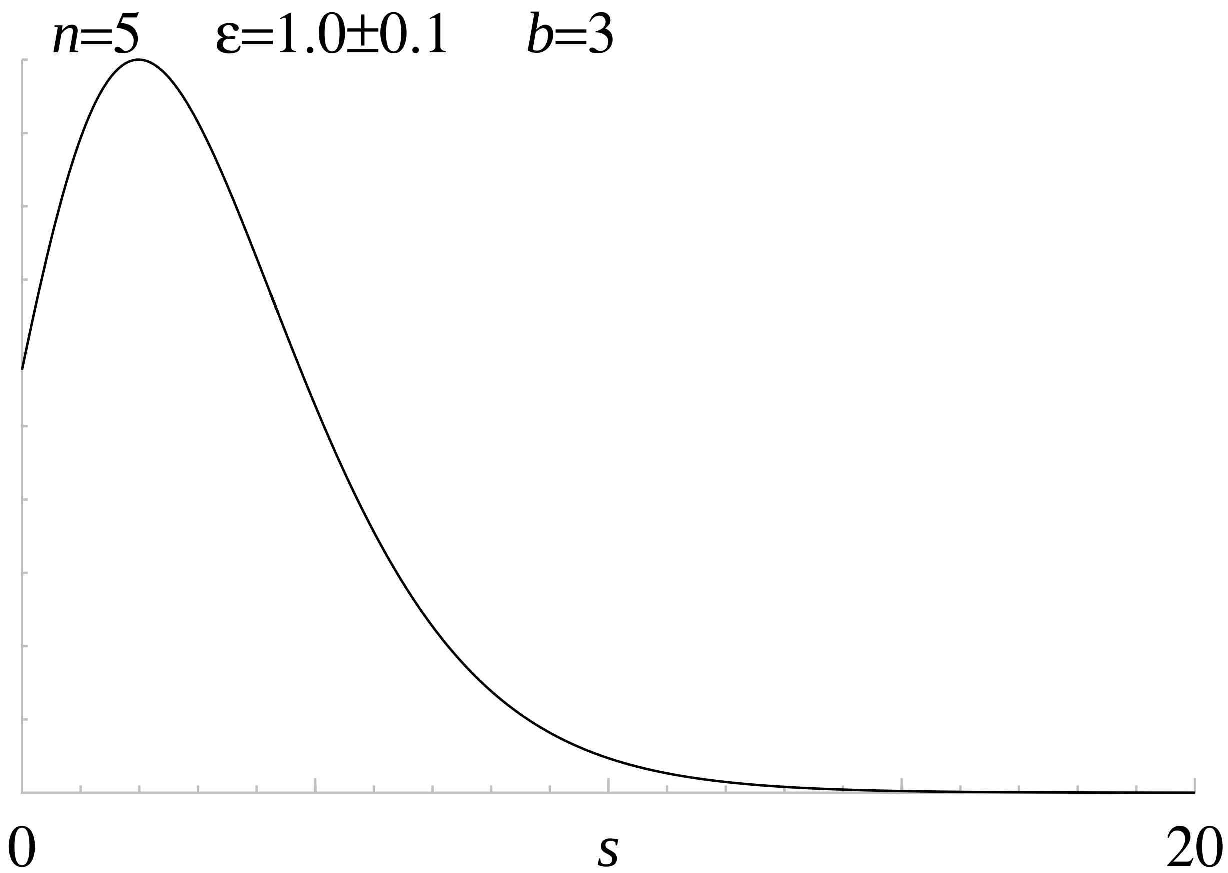
becomes the prior for ϵ in the main measurement. The mean of $p(\epsilon|m)$ is $(m+1)/\kappa$. (Calibration measurement for ϵ.)

The subsidiary measurement for b observes r events with Poisson rate ωb, where ω is a known constant. The subsidiary posterior,

$$p(b|r) = \frac{\omega(\omega b)^r e^{-\omega b}}{r!}$$

becomes the prior for b in the main measurement. The mean of $p(b|r)$ is $(r+1)/\omega$. (Sideband determination of b.)

The posterior $p(s|n)$ is calculated analytically, given in
www-cdf.fnal.gov/publications/cdf7117_bayesianlimit.pdf and
www-cdf.fnal.gov/publications/cdf7232_blimitguide.pdf.

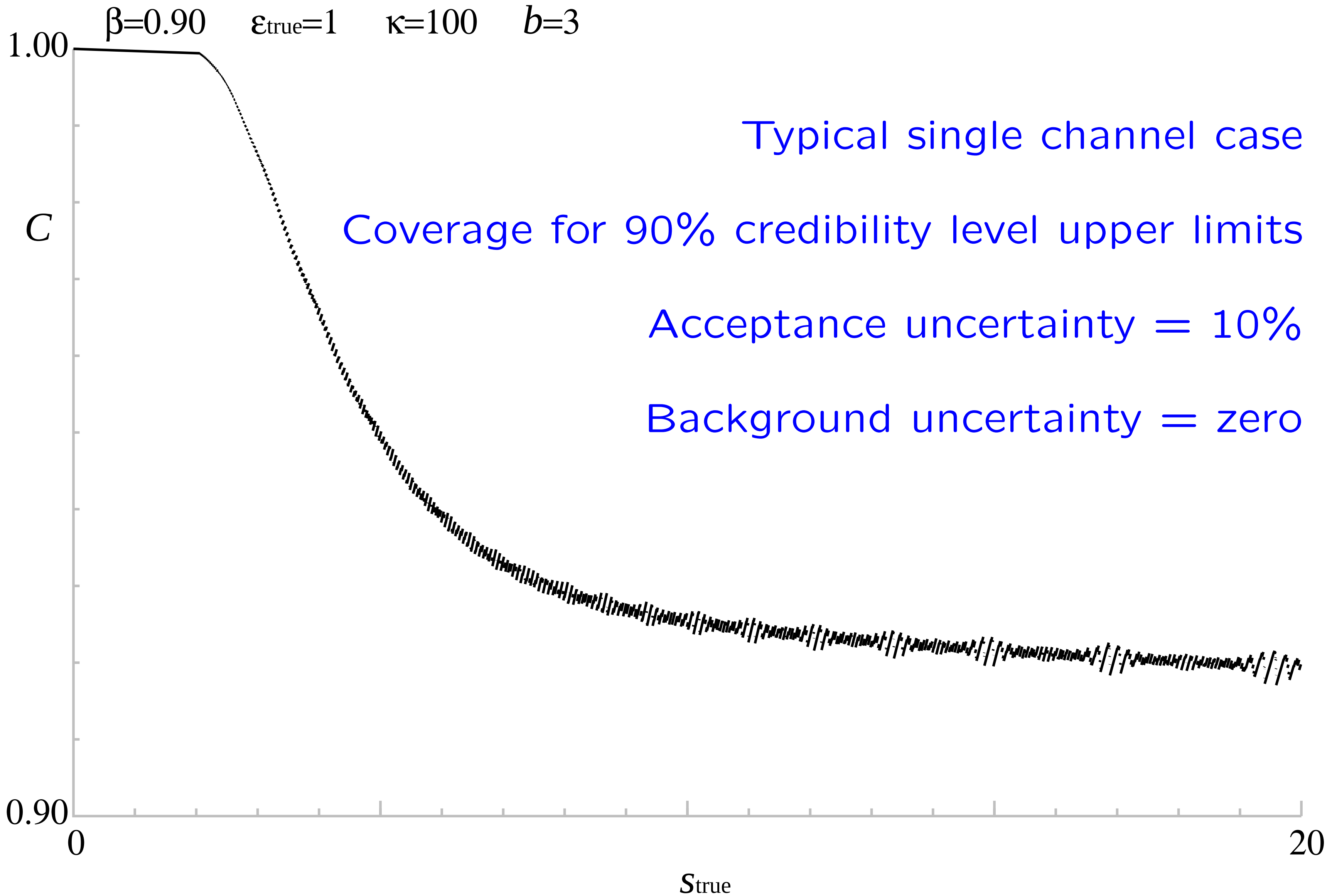


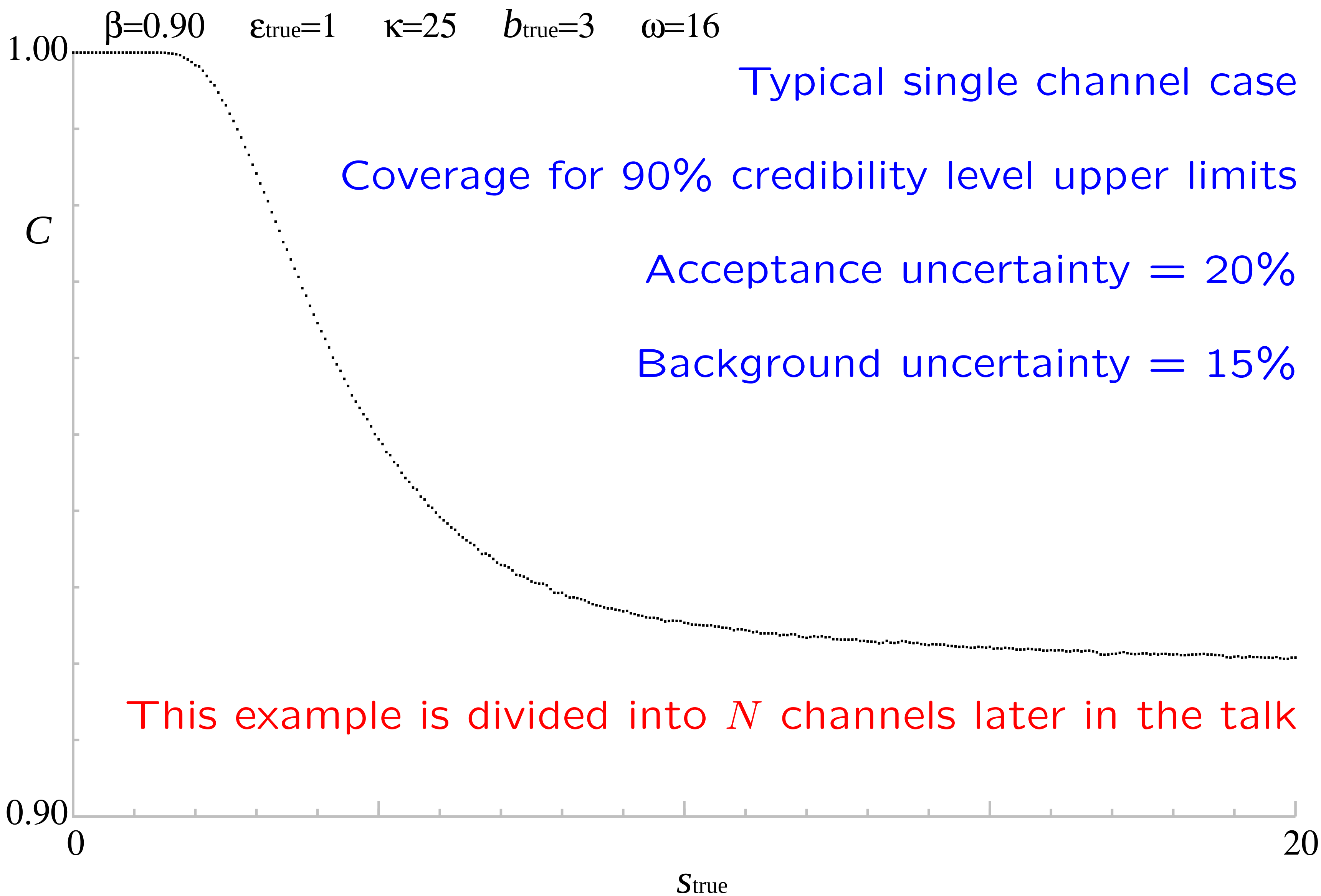
An example $p(s|n)$ with b fixed ($\kappa = 100$ and $m = 99$).

We employ objective Bayesian methodology. The priors, which are improper (and not related to personal belief), are evaluated using a frequentist technique.

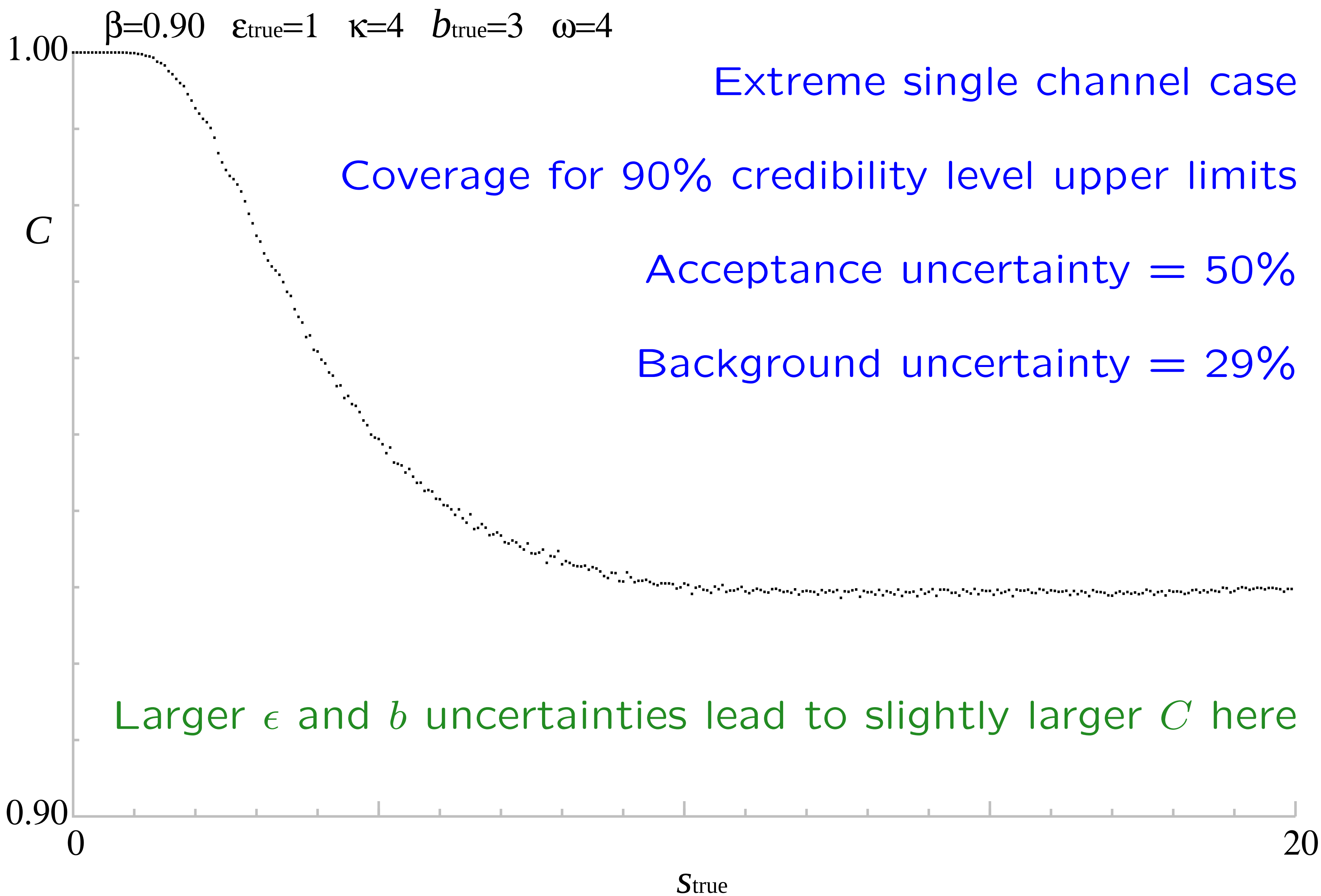
The frequentist coverage probability C is used as a diagnostic to check the performance of the limit setting scheme. For upper limits on s, C is the probability that, for fixed (true) values of the parameter of interest s and nuisance parameters ϵ and b, the resulting upper limit will be larger than s_{true}. The coverage is calculated by summing over all possible outcomes of the main and subsidiary measurements.

For this single channel case, $C > \beta$ for every combination of s_{true}, ϵ_{true}, and b_{true} tested, with this choice of priors, even when uncertainties on ϵ and b are very large. Although opinions differ on whether *any* undercoverage is acceptable, large undercoverage is considered bad. The single channel test case passes this test.


β=0.90 ε true=1 κ=100 b=3
1.00
C
0.90
0
20
s true
Typical single channel case
Coverage for 90% credibility level upper limits
Acceptance uncertainty = 10%
Background uncertainty = zero



β=0.90 εtrue=1 κ=25 btrue=3 ω=16
1.00
C
0.90
0
20
strue
Typical single channel case
Coverage for 90% credibility level upper limits
Acceptance uncertainty = 20%
Background uncertainty = 15%
This example is divided into N channels later in the talk


β=0.90 ε true=1 κ=4 b true=3 ω=4
1.00
C
0.90
0
20
s true
Extreme single channel case
Coverage for 90% credibility level upper limits
Acceptance uncertainty = 50%
Background uncertainty = 29%
Larger ϵ and b uncertainties lead to slightly larger C here

Multiple Channels

Given N channels, and n_k observed events in the kth channel, $k = 1, 2, \ldots, N$, the Poisson probability of obtaining the observed result is

$$\prod_{k=1}^{N} \frac{e^{-(s\epsilon_k + b_k)}(s\epsilon_k + b_k)^{n_k}}{n_k!}$$

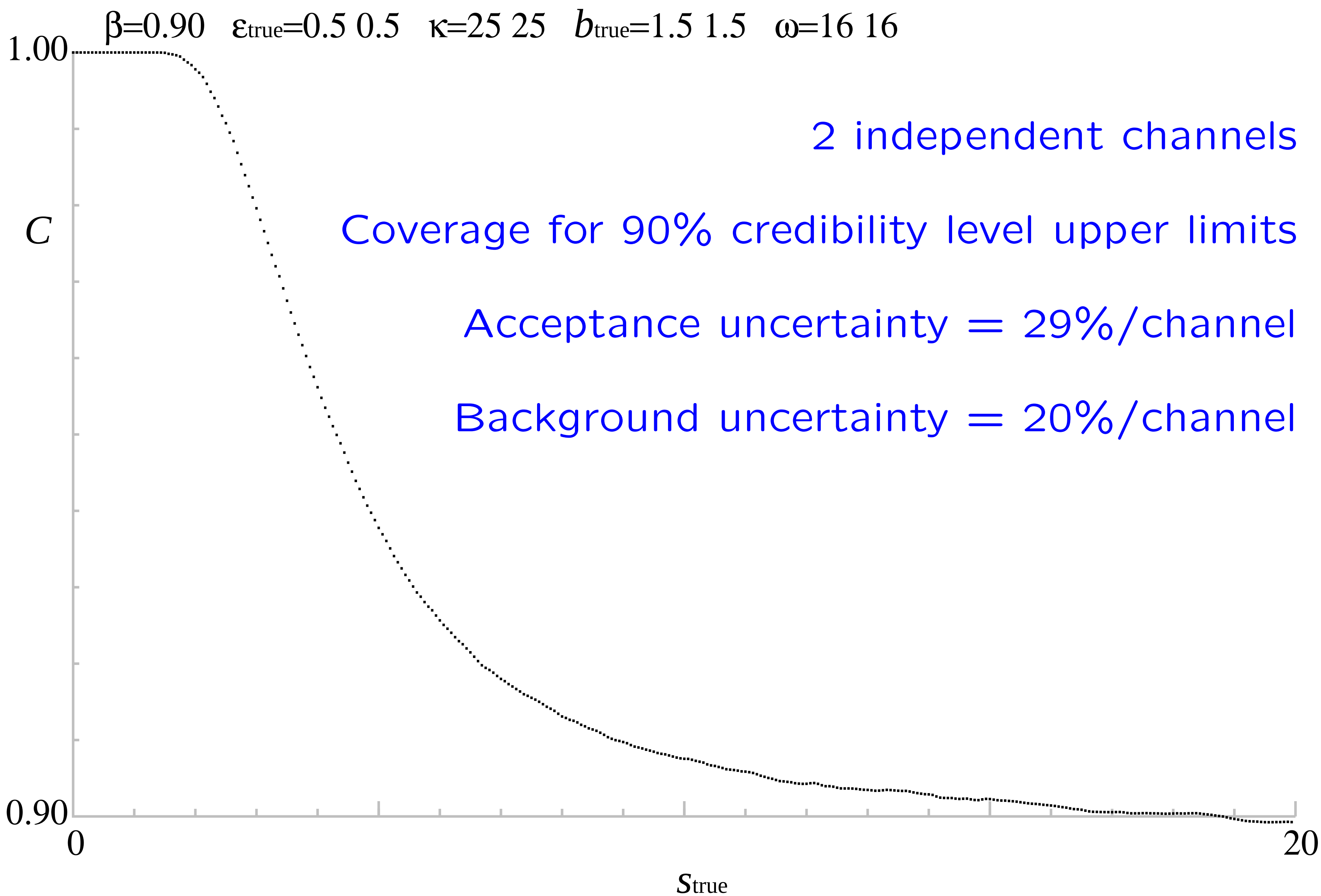
where s the cross section and ϵ_k and b_k are the acceptance and expected background for the kth channel, respectively. One multiplies by $2N$ nuisance priors and marginalizes.

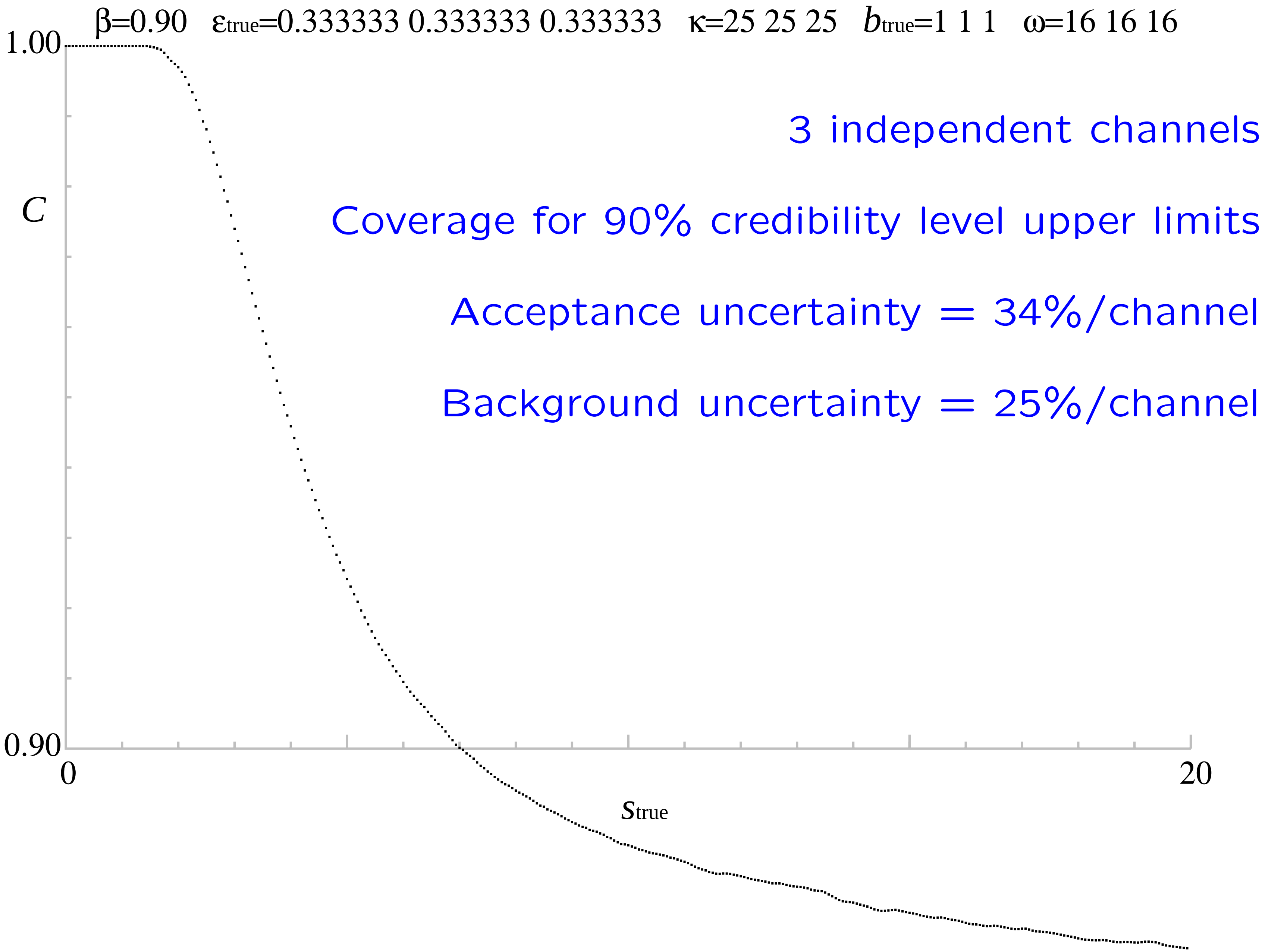
www-cdf.fnal.gov/publications/cdf7587_genlimit.pdf describes a MC integration approach to calculating the Bayesian posterior for s, given a prior flat in s, but no restrictions on the nuisance priors.

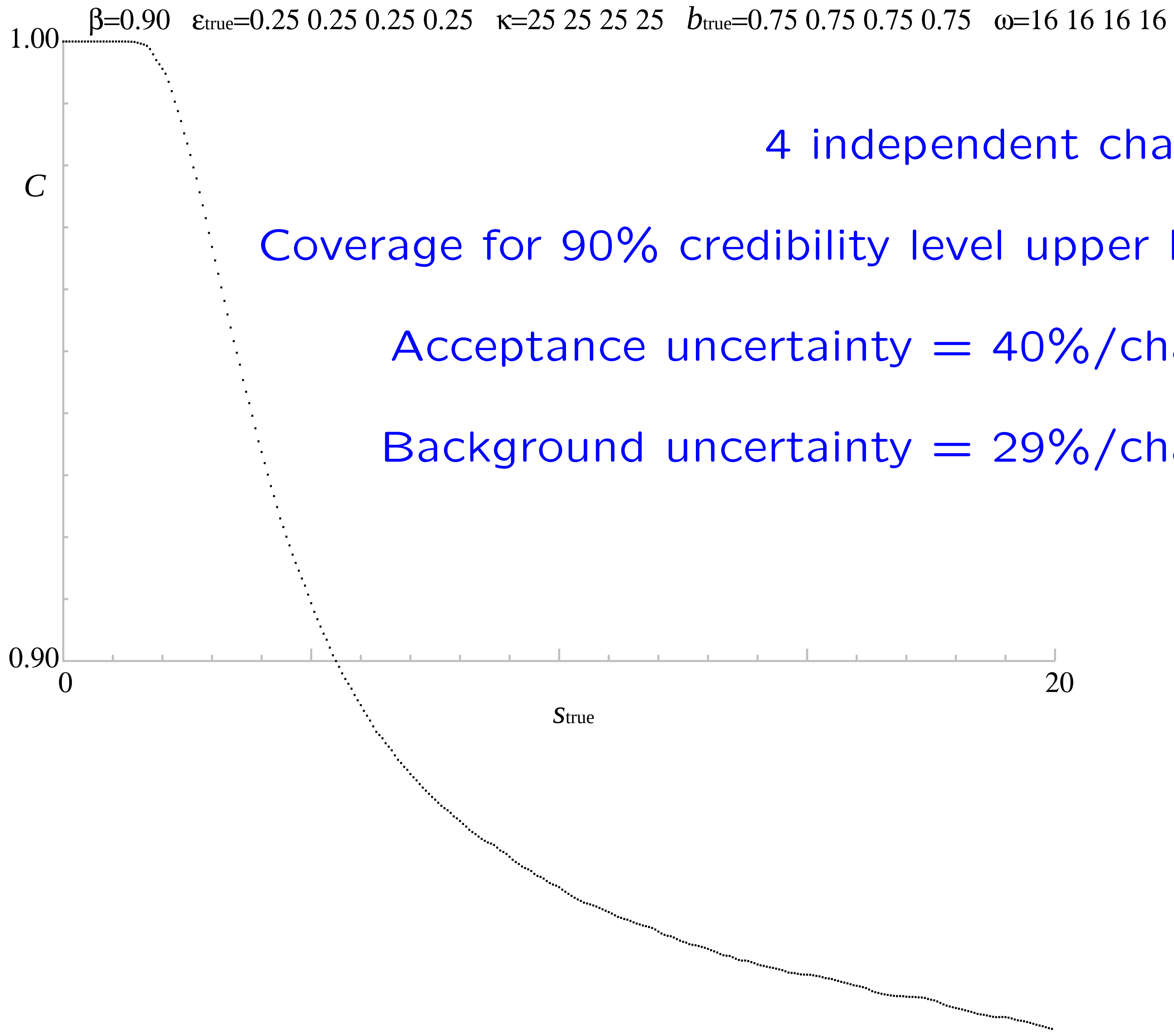
2nd test case: UL for N-independent-channel Poisson process

We specify that the data of the 1st test case (both the main measurement and the subsidiary measurements) are divided into N samples that are treated independently, to derive an upper limit on the common parameter s. Flat priors are specified for the $2N$ subsidiary measurements, leading to $2N$ subsidiary posteriors that become the nuisance priors for the main measurement. The prior for s remains flat.

For this Poisson example, we find that, when the size of the initial subsidiary data sets is not large, dividing into N independent channels drives C progressively further down as N increases.


β=0.90 ε_true=0.5 0.5 κ=25 25 b_true=1.5 1.5 ω=16 16
1.00
C
0.90
0
20
s_true
2 independent channels
Coverage for 90% credibility level upper limits
Acceptance uncertainty = 29%/channel
Background uncertainty = 20%/channel


β=0.90 ε_true=0.333333 0.333333 0.333333 κ=25 25 25 b_true=1 1 1 ω=16 16 16
1.00
C
0.90
0
20
s_true
3 independent channels
Coverage for 90% credibility level upper limits
Acceptance uncertainty = 34%/channel
Background uncertainty = 25%/channel



4 independent channels

Coverage for 90% credibility level upper limits

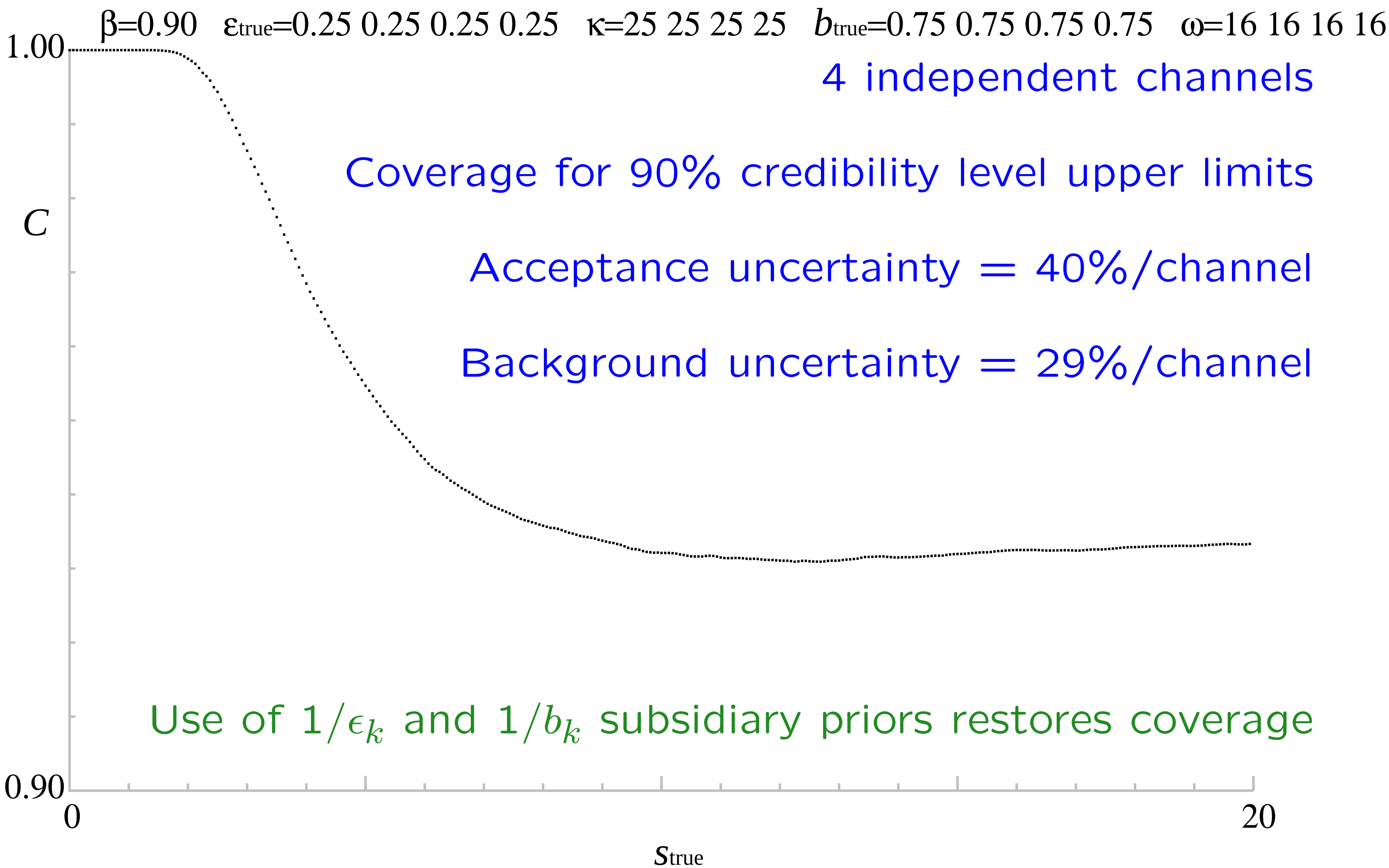
Acceptance uncertainty = 40%/channel

Background uncertainty = 29%/channel

The fault is in our choice of priors for the Poisson subsidiary measurements. E.g., a flat prior for each channel's ϵ_k subsidiary measurement yields an ϵ^{N-1} prior for the total acceptance, creating a large bias when $N > 2$. (Same bias problem for b.)

With respect to UL's, a flat prior for s leads to a bias producing overcoverage in simple Poisson cases. This bias in the subsidiary measurements leads to undercoverage in the main measurement, since an overestimate of ϵ or b leads to an underestimate for s. In our test case, using a flat prior is "conservative" for s, but "anticonservative" for ϵ and b. When $N = 1$, they roughly balance. When $N > 2$, the subsidiary priors dominate.

For our test case, a "perfect" solution is available: Use $1/\epsilon_k$ and $1/b_k$ (Jeffreys "other") priors for the subsidiary measurements.


β=0.90 ε_true=0.25 0.25 0.25 0.25 κ=25 25 25 25 b_true=0.75 0.75 0.75 0.75 ω=16 16 16 16
4 independent channels
Coverage for 90% credibility level upper limits
Acceptance uncertainty = 40%/channel
Background uncertainty = 29%/channel
Use of $1/\epsilon_k$ and $1/b_k$ subsidiary priors restores coverage
1.00
C
0.90
0
20
s_true

With this choice of subsidiary priors, the nuisance priors for the kth channel become

$$p(\epsilon_k|m_k) = \frac{\kappa_k(\kappa_k\epsilon_k)^{m_k-1}e^{-\kappa_k\epsilon_k}}{(m_k-1)!} \qquad p(b_k|r_k) = \frac{\omega_k(\omega_k b_k)^{r_k-1}e^{-\omega_k b_k}}{(r_k-1)!}$$

The means are m_k/κ_k and r_k/ω_k, respectively, eliminating the bias:

$$\langle m_k/\kappa_k \rangle = \epsilon_{\text{true},k} \qquad \langle r_k/\omega_k \rangle = b_{\text{true},k}$$

That is, the mean of the nuisance prior is now an unbiased estimator of the true value of the nuisance parameter.

progress since September 2005

Preceding pages from PHYSTAT05 talk. See:
www.physics.ox.ac.uk/phystat05/proceedings/files/heinrich.ps

Unfortunately, coverage studies like this are very time consuming (both CPU time and physicist time). We propose a method that should catch many problems before the full computation of frequentist coverage. The proposed test is as follows:

1. Compute the Bayesian limit s_N (CL=β) for a simple test case in the N-channel scheme to be tested. Must specify the outcome for the main and subsidiary measurements.
2. Make a corresponding test case with all ϵ_k and b_k proportional to one master ϵ parameter and one master b parameter. Choose constants of proportionality from the outcomes specified for the subsidiary measurements in step 1; i.e. force the nuisance parameters to be related in the same way as the step 1 subsidiary measurements' outcomes. In effect, the $2N$ subsidiary measurements are combined into a single subsidiary measurement each for ϵ and b.
3. In step 2, by forcing the nuisance parameters to be 100% correlated, we have coerced the test case of step 1 into a single channel format. Compute the posterior p.d.f. for the parameter of interest s for the step 2 test case (easy since there is now only one channel).

4. Note that the step 2 test case is fictitious, not corresponding to the reality of the step 1 case. We have added information in step 2, without altering the data, which should result in a better limit on the parameter of interest. Compute the Bayesian credibility β' of the limit s_N obtained in step 1 using the posterior from step 3. We expect that $\beta' > \beta$; if this occurs, the test is passed. Having β' significantly less than β indicates a pathological choice of subsidiary priors.

Four-channel example: observe 3 events/channel. The 8 subsidiary measurements are:

$$m_k = 6 \qquad \kappa_k = 25 \qquad r_k = 12 \qquad \omega_k = 16$$

for $k = 1, 2, 3, 4$.

Calculate the upper limit for the cross section s in the 4 channel case where we use priors flat in ϵ and b for the subsidiary measurements:

$$\text{upper limit at 90.0\% credibility level} = 14.4724 \pm 0.0040$$

Next, compute Bayesian credibility of this limit in the single channel case obtained by combining the main and subsidiary measurements into a single channel. We integrate the marginalized posterior from $s = 0$ to $s = 14.4724$, and obtain

$$\text{corresponding single channel credibility} = 0.810289$$

When combining 4 channels into 1 channel, we are adding the additional information that the acceptance and background rates are the same for each channel. Using this additional information, we expect to set a slightly better upper limit, but instead find that we would have to integrate out to significantly higher s to reach 0.9, so we get a worse limit for the single channel case.

The proposed fix to restore coverage that uses $1/\epsilon$ and $1/b$ priors for the subsidiary measurements, in contrast, yields:

$$\text{upper limit at 90.0\% credibility level} = 17.1892 \pm 0.0053$$

for the 4 channel case, and obtains

$$\text{corresponding single channel credibility} = 0.905875$$

which is consistent with expectations: adding information leads to a slightly better limit.

Conclusions

- The multichannel case involves a multidimensional nuisance prior. In hindsight, this should have led us to distrust a prior flat in multiple dimensions, since this is well known to lead to problems.

- Our example is not entirely realistic, as it specifies unusually low precision calibrations. Also, correlations among the ϵ_k and b_k, which would effectively reduce the dimensionality, are absent. But extreme cases are useful for testing the method.

- Marginalization over nuisance parameters using Bayesian priors is a common feature of many methods for setting limits. Using unbiased priors will help avoid pathologies.

Conclusions (continued)

- The $1/\epsilon_k$ and $1/b_k$ subsidiary priors are matched to this Poisson case. Other cases will require different solutions.

- In the objective Bayesian approach, the choice of subsidiary priors is just as important as the choice of prior for the parameter of interest in the main measurement. Switching to $1/\epsilon_k$ and $1/b_k$ subsidiary priors to remove the bias in the nuisance priors raised the coverage significantly, and may make use of $1/\sqrt{s}$ prior in the main measurement more appealing.

- Coverage calculations are useful in revealing poor choices of prior in the objective Bayesian approach.