Importance Sampling and Error Probabilities

Michael Woodroofe

Departments of Mathematics and Statistics

University of Michigan

July 13, 2010

Prologue

Siegmund (1975, *Ann. Stat.*)

- Importance Sampling
- Boundary Crossing Probabilities–Repeated Tests
- Clever Trick
- Dramatic Reductions

Today

- Adapt and Simplify
- MC Estimate of $p < 10^{-6}$ with 3600 replications.

Marked Poisson Variables

- g and h are mass functions (or densities), assumed known
- $N \sim \text{Poisson}(b+s)$, where $b > 0$ is known, but $s \geq 0$ is not
- $X_1, X_2, \cdots \in \mathcal{X}$ are *i.i.d.* with mass function

$$f_s = \frac{bg + sh}{b + s}$$

- Observe N and $X_1, \cdots, X_N$
- Implicitly, $N = N^b + N^s$, etc. $\cdots$.

The Likelihood Function

Clearly,

$$P_s[N = n, X_1 = x_1, \cdots, X_n = x_n]$$
$$= \frac{1}{n!}(b+s)^n e^{-(b+s)} \times \prod_{i=1}^{n} \left[\frac{bg(x_i) + sh(x_i)}{b+s}\right].$$

So, with $\mathbf{x} = (x_1, \cdots, x_n)$ and $r = h/g$,

$$L(s|n, \mathbf{x}) = \frac{1}{n!} \prod_{i=1}^{n} [bg(x_i) + sh(x_i)] e^{-(b+s)}$$
$$= L(0|n, \mathbf{x}) \prod_{i=1}^{n} \left[1 + \frac{s}{b} r(x_i)\right] e^{-s}.$$

The MLE and LRT

Let $r_i = r(x_i)$. If $r_1 + \cdots + r_n \leq b$, then the MLE is $\hat{s} = 0$.

Otherwise

$$\sum_{i=1}^{n} \frac{r_i}{b + r_i \hat{s}} = 1.$$

The LRT Statistics is

$$\lambda = \lambda(n, \mathbf{x}) = 2 \log \left[\frac{L(\hat{s}|n, \mathbf{x})}{L(0|n, \mathbf{x})} \right]$$

and, for large b,

$$P_{s=0}[\lambda > c] \approx (1 - \Phi(\sqrt{c})$$

?? Does this work for $c = 25$??

Direct Simulation

Let E denote an event, and $p = P_{s=0}(E)$. To approximate this by Monte Carlo

- Generate M pairs $(N_i, \mathbf{X}_i), i = 1, \cdots, M$
- Compute
$$\hat{p} = \frac{1}{M} \sum_{i=1}^{M} \mathbf{1}_E(N_i, \mathbf{X}_i),$$
where $\mathbf{1}_E(n, \mathbf{x}) = 1$ or 0 for $(n, \mathbf{x}) \in E$ or $(n, \mathbf{x}) \notin E$.
- For small p: Making SE 10% of p requires $M \geq 100/p$.
- For $p = 3 \times 10^{-7}$, this exceeds 3×10^{8}.

Mixtures

Let π be a density on $[0, \infty)$ and

$$g(n, \mathbf{x}) = \int_0^\infty P_s[N = n, X_1 = x_1, \cdots, X_n = x_n]\pi(s)ds.$$

Then

$$\begin{aligned} g(n, \mathbf{x}) &= \int_0^\infty L(s|n, \mathbf{x})\pi(s)ds \\ &= L(0|n, \mathbf{x}) \int_0^\infty \prod_{i=1}^{n} \left[1 + \frac{s}{b} r_i\right] e^{-s}\pi(s)ds, \end{aligned}$$

and g is a mass function; that is, $\sum_{n=0}^\infty \sum_{\mathbf{x} \in \mathcal{X}^n} g(n, \mathbf{x}) = 1$.

Some Algebra

Write

$$p = \sum_{n=0}^{\infty} \sum_{\mathbf{x} \in \mathcal{X}^n} \mathbf{1}_E(n, \mathbf{x}) L(0|n, \mathbf{x})$$

$$= \sum_{n=0}^{\infty} \sum_{\mathbf{x} \in \mathcal{X}^n} \frac{\mathbf{1}_E(n, \mathbf{x})}{\Gamma(n, \mathbf{x})} g(n, \mathbf{x})$$

where

$$\Gamma = \Gamma(n, \mathbf{x}) = \frac{g(n, \mathbf{x})}{L(0|n, \mathbf{x})}.$$

Relation Between Γ and λ

Let

$$\ell(s) = \log[L(s|N, \mathbf{X})]$$

Then

$$\Gamma = \int_0^\infty \left[\frac{L(s|n, \mathbf{x})}{L(0|n, \mathbf{x})}\right] \pi(s) ds \leq e^{\frac{1}{2}\lambda}$$

and

$$\Gamma \approx \sqrt{\frac{2\pi}{|\ell''(\hat{s})|}} e^{\frac{1}{2}\lambda} \pi(\hat{s})$$

for large N and $\hat{s} >> 0$.

Importane Sampling

- Generate $(N_i, \mathbf{X}_i),\ i = 1, \cdots, M$, from g
- Compute

$$\hat{p} = \frac{1}{M} \sum_{i=1}^{M} \frac{\mathbf{1}_E(N_i, \mathbf{X}_i)}{\Gamma(N_i, \mathbf{X}_i)}$$

  and

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^{M} \left[\frac{\mathbf{1}_E(N_i, \mathbf{X}_i)}{\Gamma(N_i, \mathbf{X}_i)} \right]^2 - \hat{p}^2.$$

- Report

$$\hat{p} \pm \frac{\hat{\sigma}}{\sqrt{M}}.$$

Some Details

To generate $(N, \mathbf{X})$ from g

- Generate s from π
- Generate N from Poisson$(b + s)$
- Generate $X_1, \cdots, X_N$ from f_s

Example

- $b = 10$, $\mathcal{X} = [0, 1]$, $g(x) = 1$, and $h(x) = 2x$.
- $E = \{\lambda > 25\}$ and $M = 3600$
- $\hat{p} = (2.3964 \cdots) \times 10^{-7} \pm (2.3225 \cdots) \times 10^{-8}$
- Nominal (asymptotic) $p = (2.866 \cdots) \times 10^{-7}$

Remarks

- Smaller M; more complicated algorithm.
- Requires
$$\sum_{(n,\mathbf{x})\in E} \frac{g(n,\mathbf{x})}{\Gamma^2(n,\mathbf{x})} < \infty.$$
- π is not a prior
- Want
$$\int_0^\infty (1+s)\pi(s)ds < \infty.$$
- Used $\pi(s) = b^{-1}e^{-s/b}$ in the Example; too sharp?
- Γ as a test statistic; same technique can be used.

More Remarks

- Unknown b, with some effort
- Unknown g and/or h (parametric)?
- Let λ_t the the log LRS after t time units, $t = 1, \cdots, T$. Then the same techniques can be used on

$$P_{s=0}\left[\max_{t \leq T} \lambda_t > c\right].$$