Large Deviations

Will Perkins

March 5, 2013

Probabilities of Rare Events

Let X be the number of heads in n flips of a fair coin. The central limit theorem tells us something about the probability that X is large:

$$\Pr[X \geq \frac{n}{2} + \sqrt{n}] \to .025$$

What does the CLT have to say about

$$\Pr[X \geq \frac{n}{2} + n/3]$$

?

$\Pr \to 0$ as $n \to \infty$.

Asymptotics

Using Stirling's Formula, we know something much more precise:

$$\begin{aligned}
\Pr[X \geq \frac{n}{2} + n/3] &= \sum_{k=5n/6}^{n} \binom{n}{k} 2^{-n} \\
&\leq \frac{n}{6} \binom{n}{5n/6} 2^{-n} \\
&\sim \frac{n}{6} \frac{n^n \sqrt{2\pi n} 2^{-n}}{(5n/6)^{5n/6} (n/6)^{n/6} 2\pi \sqrt{5/36}} \\
&= e^{n\left[-\frac{5}{6}\log(5/6) - \frac{1}{6}\log(1/6) - \log 2 + o(1)\right]} \\
&= e^{-.2426n + o(n)}
\end{aligned}$$

Asymptotics

We also have a *lower bound*:

$$\begin{aligned}\Pr[X \geq \frac{n}{2} + n/3] &\geq \Pr[X = \frac{n}{2} + n/3] \\ &= \binom{n}{5n/6} 2^{-n} \\ &= e^{-.2426n + o(n)}\end{aligned}$$

Notice that this is much more precise than saying the probability tends to 0 with n.
And in fact the CLT would be 'wrong':

$$\Pr[Z \geq 2\sqrt{n}/3] \approx e^{-2n/9} \approx e^{-.222n}$$

Large Deviations

This basic example already gives us some of the themes of Large Deviation Theory:

1. Find the asymptotic probabilities of rare events - how do these decay to 0 as $n \to \infty$?
2. 'Large' deviations are those past the point at which the CLT gives the right answer.
3. In general these probabilities will be exponentially small in n: e^{-cn} for some c. We are concerned with finding the correct c. I.e., find:

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr[A_n^{rare}]$$

4. This involves an upper and a lower bound.

Rate Functions

Let $X_1, X_2, \ldots X_n$ be iid random variables and define the event $A_n(x)$ as:

$$A_n(x) = \left\{ \frac{X_1 + \cdots + X_n}{n} \geq x \right\}$$

The rate function of the random variable X_i is:

$$I(x) := -\lim_{n \to \infty} \frac{1}{n} \log \Pr[A_n(x)]$$

Rate Functions

We've seen two rate functions already:

- $X_i \sim Ber(1/2)$.

  $$I(x) = x \log x + (1-x)\log(1-x) + \log 2$$

  for $x \in (1/2, 1)$.
- $X_i \sim N(0,1)$. In this case $\frac{X_1+\cdots+X_n}{n} \sim N(0, 1/n)$, and so HW 1, Problem 5 says:

  $$I(x) = x^2/2$$

  for $x > 0$.

[After normalizing properly, compare the two!]

Chernoff Bounds

Coin flips and normal random variables are things we know very well. What about a more general statement?

"Chernoff Bounds" are large-deviation upper bounds for random variables satisfying various conditions.

We want a statement like:

$$\Pr\left[\frac{X_1 + \cdots + X_n}{n} > a\right] \leq e^{-g(a)n}$$

for some function $g(a)$ that depends on our hypotheses about our iid mean 0 random variables X_i.

A Chernoff Bound for Bounded RV's

Let $X_1, \ldots$ be iid RV's with mean 0, and assume that $a \leq X_i \leq b$ with probability 1. We will prove that

$$\Pr\left[X_1 + \cdots + X_n > t\right] \leq e^{-\frac{2t^2}{(b-a)^2}}$$

[compare with previous by setting $t = an$]

The proof technique is a good one to know since it can be adapted to many different situations.

Proof

Let $Y = X_1 + \cdots + X_n$.

Step 1: An 'Exponential' Markov's Inequality:

$$\Pr[Y \geq t] \leq \frac{\mathbb{E}e^{\lambda Y}}{e^{\lambda t}}$$

for any $\lambda > 0$. [See HW 3, # 5]

Proof

Step 2: An upper bound on $\mathbb{E}e^{\lambda Y}$.

$$\begin{aligned}\mathbb{E}e^{\lambda Y} &= \left(\mathbb{E}e^{\lambda X_i}\right)^n \text{ (independence)}\\ &= \left(e^{\lambda^2(b-a)^2/8}\right)^n \text{ (uses convexity)}\end{aligned}$$

Proof

Step 3: Optimize over λ with calculus.

$$\Pr[Y \geq t] \leq e^{\lambda^2(b-a)^2/8-\lambda t}$$

Minimize $\lambda^2(b-a)^2/8 - \lambda t$ over choices of λ:

$$\lambda = \frac{4t}{(b-a)^2}$$

Which gives

$$\Pr[Y \geq t] \leq e^{-2t^2/(b-a)^2}$$

Chernoff Bounds

Notice that the upper bounds we get with this method are much much better than we would get with Chebyshev's inequality.

Show that in $G(n, 1/2)$

General Large Deviation Theory

We can replicate the above proof for a general random variable X_i with distribution μ and mean m. Assume we know the moment generating function, $M(\lambda) = \mathbb{E}e^{\lambda X_i} = \int e^{\lambda x}\, d\mu$.

Upper Bound:
Let $S_n = X_1 + \cdots + X_n$.

$$\Pr[S_n \geq tn] \leq e^{-\lambda nt}[M(\lambda)]^n$$

Let $\psi(\lambda) = \log M(\lambda)$,

$$\frac{1}{n}\log \Pr[S_n \geq tn] \leq -\lambda t + \psi(\lambda)$$

and optimizing over λ gives

$$\frac{1}{n}\log \Pr[S_n \geq tn] \leq -\sup_{\lambda} \lambda t - \psi(\lambda)$$

General Large Deviation Theory

Lower Bound:
The lower bound comes from a procedure called "Tilting the Measure".
As we did to prove the Chernoff bound, we find the maximizer λ_0 of $\lambda t - \psi(\lambda)$ by setting the derivative to 0:

$$t = \psi'(\lambda) = \frac{M(\lambda_0)'}{M(\lambda_0)} = \frac{1}{M(\lambda_0)} \int x e^{\lambda_0 x} \, d\mu(x)$$

This leads to the idea of replacing μ with the distribution $\mu_0 = \frac{e^{\lambda_0 x}}{M(\lambda_0)} \mu$

General Large Deviation Theory

From the above, we know that $\int x \, d\mu_0(x) = t$; that is, we have shifted the mean of the distribution to t.

Let $A = \{\frac{X_1 + \cdots + X_n}{n} \geq t\}$ and $A_\delta = \frac{X_1 + \cdots + X_n}{n} \in [t, t + \delta]$.

$$\Pr_{\mu^{(n)}}(A) = \int_A [M(\lambda_0) e^{-\lambda_0 x}]^n \, d\mu_0^{(n)}(x)$$

and,

$$\mu^{(n)}(A) \geq \mu^{(n)}(A_\delta) \geq [M(\lambda_0) e^{-\lambda_0 (t+\delta)}]^n \mu_0^{(n)}(A_\delta)$$

General Large Deviation Theory

Now we can use the CLT for $\mu_0^{(n)}$:

$$\mu_0^{(n)}(A_\delta) \to \frac{1}{2}$$

Then taking logarithms and dividing by n, we have:

$$\lim_{n\to\infty} \frac{1}{n} \log \mu^{(n)}(A) \geq \psi(\lambda_0) - \lambda_0(t+\delta)$$

But since this is true for every $\delta > 0$, we can take $\delta \to 0$ and get our lower bound:

$$\lim_{n\to\infty} \frac{1}{n} \log \mu^{(n)}(A) \geq \psi(\lambda_0) - \lambda_0 t$$