1 Hoeffding's Inequality

If $X_1 ... X_n$ are independent, bounded random variables with $a_i < X_i < b_i$ and means $\mu_1 ... \mu_n$, and let $\mu = \frac{1}{n}\sum_{i=1}^{n} \mu_i$

$$P(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu] \geq \epsilon) \leq \exp \frac{-2n^2\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2} \tag{1}$$

2 Bernstein's Inequality

$X_1 ... X_N$ are independent, bounded random variables with $0 \leq X_i \leq E[X_i] + b$ Let $V = \sum [V[X_i]$

$$P(\sum X_i - \sum E[X_i] \geq \epsilon) \leq e^{-\frac{\epsilon^2}{2\sum V[X_i] + 2b\epsilon/3}} \implies P\left(\cdot \geq \frac{2b\log(1/\delta)}{3}) + \sqrt{2\log(1/\delta)\sum V[X_i]}\right) \leq \delta \tag{2}$$

$$P(\frac{\sum X_i}{n} - \frac{\sum E[X_i]}{n} \geq \epsilon) \leq e^{-\frac{n\epsilon^2}{2\frac{1}{n}\sum V[X_i] + 2b\epsilon/3}} \implies P\left(\cdot \geq \frac{2b\log(1/\delta)}{3n}) + \frac{\sqrt{2\log(1/\delta)\sum V[X_i]}}{n}\right) \leq \delta \tag{3}$$

Estimating bounds for sums of bernoulli RV's with random denominator

Suppose we have samples (X_t, Y_t) for $t = 1...n$ drawn from the joint distribution $P(X, Y)$, where X and Y are binary random variables and $P(X) = p$ and we want to estimate $\mu = P(Y = 1|X = 1)$

Importance sampling estimate

Let:

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} Z_t, \text{ where } Z_t = \frac{Y_t \mathbb{1}\{X_t = 1\}}{p} \in \{0, \frac{1}{p}\} \tag{4}$$

$$W_t = pZ_t,\ W_t \in \{0, 1\},\ V[W_t] \leq p \tag{5}$$

$$\text{Hoeff.} \implies P(\hat{\mu} - \mu \geq \epsilon) \leq e^{-2np^2\epsilon^2} \implies P(\hat{\mu} - \mu \geq \frac{1}{p}\sqrt{\frac{1}{2n}\log\frac{1}{\delta}}) \leq \delta \tag{6}$$

$$\text{Berns.} \implies P(\mu - \hat{\mu} \geq \epsilon) \leq e^{-\frac{np\epsilon^2}{2(1+\epsilon/3)}} \implies P(\mu - \hat{\mu} \geq \frac{2\log(1/\delta)}{3np} + \sqrt{\frac{2}{np}\log\frac{1}{\delta}}) \leq \delta \tag{7}$$

Standard Estimator

Let

$$\hat{\mu} = \frac{\sum_{t=1}^{n} Y_t \mathbb{1}\{X_t = 1\}}{\sum_{t=1}^{n} \mathbb{1}\{X_t = 1\}} \tag{8}$$

We have a problem if we don't have a lower bound on $\sum_{t=1}^{n} \mathbb{1}\{X_t = 1\}$ as the estimator may be undefined. Assume $s = \sum_{t=1}^{n} \mathbb{1}\{X_t = 1\} > 0$, then:

$$\text{Hoeff.} \implies P(\hat{\mu} - \mu \geq \epsilon) \leq e^{-2s\epsilon^2} \implies P\left(\hat{\mu} - \mu \geq \sqrt{\frac{1}{2s}\log(\frac{1}{\delta})}\right) \tag{9}$$

3 The importance sampling estimator

If we know the probability that $X = 1$ we can instead use importance sampling.

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} \frac{Y_t \mathbb{1}\{X_t = 1\}}{P(X_t = 1)} \tag{10}$$

The estimator is unbiased:

$$\begin{aligned}
E[\hat{\mu}] =& \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{Y_t \mathbb{1}\{X_t = 1\}}{P(X_t = 1)}\right] = \frac{1}{np}\sum_{t=1}^{n} E\left[Y_t \mathbb{1}\{X_t = 1\}\right] & (11)\\
=& \frac{1}{n}\sum_{t=1}^{n} \frac{(0 * P(Y_t = 0 \text{ or } X_t = 0) + 1 * P(Y_t = 1 \text{ and } X_t = 1))}{P(X_t = 1)} & (12)\\
=& \frac{1}{n}\sum_{i=1}^{n} \frac{P(Y = 1, X = 1)}{P(X = 1)} = \frac{1}{n}\sum_{i=1}^{n} P(Y = 1 | X = 1) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu & (13)
\end{aligned}$$

We can use Hoeffding's inequality to get a bound on how far the estimator is likely to be from the true value. Let:

$$Z_t = \frac{Y_t \mathbb{1}\{X_t = 1\}}{p} \in \{0, \frac{1}{p}\} \tag{14}$$

Hoeffding's inequality: If $X_1 ... X_n$ are independent observations such that $a_i < X_i < b_i$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ then:

$$P(\bar{X} - E[\bar{X}] \geq \epsilon) \leq \exp\frac{-2n^2\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2} \tag{15}$$

In our case this gives:

$$P(\frac{1}{n}\sum_{i=1}^{n} Z_t - \mu \geq \epsilon) \leq \exp\left(-2n\epsilon^2 p^2\right) \implies P(\frac{1}{n}\sum_{i=1}^{n} Z_t - \mu \geq \frac{1}{p}\sqrt{\frac{1}{2n}\log\frac{1}{\delta}}) \leq \delta \tag{16}$$

This is not so good because $\frac{1}{p}$ can be very large if p is small and its outside the log - so the bounds will grow quickly as p gets small.

We can get a tighter bound by using an Chernoff's inequality that takes account of the variance of X. Let $W_t = \mathbb{1}\{Y_t = 1, X_t = 1\} = pZ_t$, then:

$$P(\frac{1}{n}\sum_{i=1}^{n} Z_t - \mu \geq \epsilon) = P(\frac{1}{n}\sum_{i=1}^{n} pZ_t - p\mu \geq p\epsilon) = P(\frac{1}{n}\sum_{i=1}^{n} W_t - p\mu \geq p\epsilon) \tag{17}$$

Now W_t is a bernoulli random variable so:

$$V[W_t] = P(W_t = 1)P(W_t = 0) \leq P(W_t = 1) = P(Y_t = 1, X_t = 1) \leq P(X_t = 1) = p \tag{18}$$

Now do the same for the standard estimator