Math 161 0 - Probability, Fall Semester 2012-2013
Dan Abramovich

Expected Inequalities

Question: Suppose $Y(\omega) \geq \epsilon$ for all ω. What can you say about $E(Y)$?

Answer:

$E(Y) = \int_\Omega Y(\omega) f(\omega) d\omega$
$\geq \int_\Omega \epsilon f(\omega) d\omega = \epsilon.$

$\boxed{E(Y) \geq \epsilon}$

Question: Suppose $Y(\omega) \geq 0$ for all ω and $Y(\omega) > \epsilon$ for all $\omega \in F \subset \Omega$. What can you say about $E(Y)$?

Answer:

$E(Y) = \int_\Omega Y(\omega) f(\omega) d\omega$
$\geq \int_F \epsilon f(\omega) d\omega + \int_{\tilde{F}} 0 f(\omega) d\omega = \epsilon P(F).$

$\boxed{E(Y) \geq \epsilon P(F)}$

Given Ω, X, μ, σ.

Question: Write $F_\epsilon = \{\omega : |X - \mu| \geq \epsilon$. What can you say about $V(X) = E((X-\mu)^2)$?

Answer: $\boxed{V(X) \geq P(F_\epsilon)\epsilon^2}$

Chebyshev's inequality:

$$\boxed{P(|X-\mu| \geq \epsilon) \quad \leq \quad \frac{V(X)}{\epsilon^2}}$$

X_i independent, μ, σ. Write $X = A_n$, which has $E(A_n) = \mu, V(A_n) = \sigma^2/n$.

Application:

$$\boxed{P(|A_n - \mu| \geq \epsilon) \quad \leq \quad \frac{\sigma^2}{n\epsilon^2}}$$

Law of large numbers:

$$P(|A_n - \mu| \geq \epsilon) \xrightarrow[n \to \infty]{} 0$$

Law of Averages:

$$P(|A_n - \mu| \leq \epsilon) \xrightarrow[n \to \infty]{} 1$$

Note: Chebyshev actually gives an extimate! It is a blunt instrument.

Toss 100 coins. X = number of heads. X_i have $\mu = 1/2$, $\sigma = 1/2$.

We know that S_{100} is close to 50 with high probability. How high?

$P(|S_{100} - 50| > 10) = P(|A_n - .5| > .1) \leq (1/4)/(100 \times 0.1^2) = 1/4$.

You can calculate the actual results,

```
In[10]:= 1-Sum[Binomial[100,k],
                    {k,40,60}]/2.^100

Out[10]= 0.0352002
```

which are a quite bit better, so $P(|S_{100} - 50| \leq 10) = 0.9648$.

The central limit theorem is much more precise.

$S_n^* = (S_n - n\mu)/(\sqrt{n}\sigma)$.

Here $n = 100$, so

$S_{100}^* = (S_n - 50)/5$

$P(|S_\mu - 50| \leq 10) = P(|S_n^*| \leq \frac{10}{5})$

If the central limit theorem is right then this is approximately

$P(|S_n^*| \leq 2)$

$\sim P(-2 < N_{0,1} < 2)$:

```
In[3]:=
1/Sqrt[2*Pi]
 *Integrate[Exp[-x^2/2.],{x,-2,2}]

Out[3]= 0.9545
```

slightly more realistic: $P(|S_\mu - 50| \leq 10.5) = P(|S_n^*| \leq \frac{10.5}{5})$

```
In[4]:=
1/Sqrt[2.*Pi]
*Integrate[Exp[-x^2/2],{x,-2.1,2.1}]

Out[4]= 0.964271
```

(Like LLN, the CLT is not stated with an estimate for the error, but one can devise such estimates.)

In general for independent Bernoulli trials:

$\mu = p, \sigma = \sqrt{pq}, S_n^* = \frac{S_n - n\mu}{\sqrt{npq}}$

If $0 \leq a \leq b \leq n$ are integers, we want to estimate

$P(a \leq Binomial(n,p,q) \leq b)$, namely $P(a \leq S_n \leq b)$.

If we write $a^* = \frac{a-np-1/2}{\sqrt{npq}}, b^* = \frac{b-np+1/2}{\sqrt{npq}}$ then this is the same as $P(a^* < S_n^* < b^*)$ which is approximated by

$\frac{1}{\sqrt{2\pi}} \int_{a^*}^{b^*} e^{-x^2/2} dx = \int_{a^*}^{b^*} \phi(x) dx.$

We can prove this specific case of CLT, if we accept Stirling's formula. The form we prove is the following: we want to compare the contribution of $b(n,p,k)$ is about $\int_{a^*}^{b^*}\phi(x)dx$ where $a^* = \frac{k-np-1/2}{\sqrt{npq}}, b^* = \frac{k-np+1/2}{\sqrt{npq}}$.

When n is large the fundalental theorem of calculus says that

$\int_{a^*}^{b^*}\phi(x)dx \sim \phi(k^*)\Delta x$

where $k^* = \frac{k-np}{\sqrt{npq}}$ and $\Delta x = \frac{1}{\sqrt{npq}}$

It suffices to show that

$\phi(x) \sim \sqrt{npq} \cdot b(n,p,k)$.

There is a pretty reasonable computation in case $k = np$ in the book, where the result should be $1/\sqrt{2\pi}$.

$$\sqrt{npq} \cdot b(n, p, k)$$
$$\sim \frac{\sqrt{npq}p^{np}q^{nq}\sqrt{2\pi n}n^n/e^n}{(\sqrt{2\pi np}(np)^{np}/e^{np})(\sqrt{2\pi nq}(nq)^{nq}/e^{nq})}.$$

And indeed there is a magical cancellation.

Read: POLLING

Read: Genetics