Beginning of Sept.92022

Consider a geometric distribution with parameter p:

- $\mathbb{P}(X = n) = (1-p)^{n-1}p$.
- $\mathbb{E}X = 1/p$.
- $\mathbb{E}(X(X-1)) = \sum_{n=1}^{\infty} n(n-1)(1-p)^{n-2}(1-p)p = \frac{2-2p}{p^2}$, so
- $\mathrm{var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{2-2p}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1}{p^2} - \frac{1}{p}$.

Example: The coupon collector's problem. Suppose each cereal box has one of the n coupons equally likely. Let T_n be the time to get all n.

Let R be repeats and N be new coupons. The outcome is a sequence of R's and N's. Let $X_{n,k}$ be the time from getting the $(k-1)^{\text{th}}$ coupon to the k^{th} new coupon. It follows immediately that the $X_{n,k}$'s are independent from each other, with $T_n = \sum_{k=1}^{n} X_{n,k}$. In particular,

$$X_{n,k} \sim \text{geometric}\left(\frac{n-k+1}{n}\right),$$

so

$$\mathbb{E}T_n = 1 + \frac{n}{n-1} + \frac{n}{n-2} + \ldots + n = n(1 + 1/2 + \ldots + 1/n) \sim n \log n.$$

On the other hand,

$$\mathrm{var}(T_n) = \sum_{k=1}^{n} \mathrm{var}(X_{n,k}) \leqslant \sum_{k=1}^{n} \left(\frac{n}{n-k+1}\right)^2 = \frac{n^2\pi^2}{6}.$$

Since $\mathrm{var}(T_n)/(n\log n)^2 \to 0$, by D2.2.6, $(T_n - \mathbb{E}T_n)/(n \log n) \to 0$ in probability, i.e.,

$$\frac{T_n}{n \log n} \to 1 \qquad \text{in probability.}$$

0.1 Triangular Arrays

Consider a **triangular array** $\{X_{n,k} : n \geqslant k, k \leqslant k_n\}$ where the n^{th} row has k_n variables.

Theorem: D2.2.11, WLLN for triangular arrays

Let $\{X_{n,k}\}$ be given. Let $b_n \to \infty$ and

$$a_n := \sum_{k=1}^{k_n} \mathbb{E}(X_{n,k} 1_{\{|X_{n,k}| \leqslant b_n\}}).$$

Assume

$$\sum_{k=1}^{k_n} \mathbb{P}(|X_{n,k}| > b_n) \to 0 \qquad \text{as } n \to \infty \tag{1}$$

and

$$b_n^{-2}\mathbb{E}(X_{n,k}^2 1_{\{|X_{n,k}|\leqslant b_n\}}) \to 0 \qquad \text{as } n \to \infty, \tag{2}$$

then $(S_n - a_n)/b_n$ converges to 0 in probability.

In the i.i.d. case, where $X_{n,k} = X_k$ and $k_n = b_n = n$, (1) says $n\mathbb{P}(|X_1| > n) \to 0$ and (2) says $n^{-1}\mathbb{E}(X_1^2 1_{\{|X_1|\leqslant n\}}) \to 0$.

Theorem: D2.2.14, Finite mean of WLLN

Let $X_1, X_2, \ldots$ be i.i.d. with $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}X_1 = \mu$. Then $S_n/n \to \mu$ in probability without any assumption on the second moment.

Proof. We use WLLN 2.2.12. Let $\mu_n := \mathbb{E}(X_1 1_{\{|X_1|\leqslant n\}})$. We know $\mu_n \to \mu$ by DCT. Also,

$$x\mathbb{P}(|X_1| > x) = \mathbb{E}(x 1_{\{|X_1|>x\}}) = \mathbb{E}(|X_1| 1_{\{|X_1|>x\}}) \to 0$$

again using DCT. Therefore by 2.2.12 $S_n/n - \mu_n \to 0$ in probability, so $S_n/n \to \mu$ in probability. □

If $X_1 \geqslant 0, \mathbb{E}X_1 = \infty$, we can compare X_1 with the truncated variables to see $S_n/n \to \infty$. Nevertheless, we can still ask if there exist a_n, b_n such that $(S_n - a_n)/b_n \to 0$ in probability.

Example: D2.2.16 St. Petersburg paradox. Game: win 2^j if first heads toss is trial j, $j \geqslant 1$. Note that $S_n/n \to \mu$ implies that μ is the "fair price" to pay to play one game. Let X_k be the r.v. describing the among of games won by game k. Then

$$\mathbb{E}X_1 = \sum_{j\geqslant 1} 2^j 2^{-j} = o.$$

Then a_n is the "fair price for n games." By 2.2.11 (triangular array WLLN), we take $X_{n,k} = X_k$ for $k \leqslant n$ and $\{b_n\}$ to be determined. Let

$$a_n = n\mathbb{E}(X_1 1_{\{X_1 \leqslant b_n\}}).$$

We want b_n to satisfy two things:

- the truncation probability $n\mathbb{P}(X_1 > b_n) \to 0$,
- $b_n^{-2} n\mathbb{E}(X_1^2 1_{\{X_1\leqslant b_n\}}) \to 0$, and
- $b_n \leqslant c a_n$.

For tails:

$$\mathbb{P}(X_1 \geqslant 2^m) = \mathbb{P}(\text{first } m-1 \text{ all tails}) = 2^{-m+1}.$$