Distribution	MGF
Bernoulli(p)	$pe^t + 1 - p$
Geo(p)	$\frac{pe^t}{1-(1-p)e^t}$ for $t < -\log(1-p)$
Binomial(n, p)	$(pe^t + 1 - p)^n$
Poisson(λ)	$e^{\lambda(e^t - 1)}$
$U(a, b)$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$ for $t \neq 0$, 1 for $t = 0$
$N(\mu, \sigma^2)$	$e^{t\mu + \sigma^2 t^2/2}$
Gamma(α, β)	$(1 - \beta t)^{-\alpha}$ for $t < 1/\beta$
Exp(λ)	$\frac{\lambda}{\lambda - t}$ for $t < \lambda$

If $Y = aX + b,\ a, b \in \mathbb{R}$, then

$$M_Y(t) = e^{tb} M_X(at)$$

Goal: Find PDF of $Y = g(X)$

- Write down PDF f_X
- Determine $A = \{x \in \mathbb{R}: f_X(x) \neq 0\}$
- Verify that g is strictly monotone on A and find g^{-1}
- Determine the image of g, i.e., $\{g(x): x \in A\}$
- $f_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\right)\left|\frac{dg^{-1}(y)}{dy}\right| & \text{for } y \text{ in image of } g \\ 0 & \text{otherwise} \end{cases}$

Likelihood function:

$L(x_1, \ldots, x_{1000}|p) = \prod_{i=1}^{1000} f(x_i|p)$

$= f(1)^{13}\, f(0)^{987} = p^{13}(1-p)^{987}$

$$P\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq x\right) \approx \Phi(x)$$

The likelihood function for this sample is

$$L(\theta) = \prod_{i=1}^{n} f(X_i) = 2^{-n}\theta^{3n}\left(\prod_{i=1}^{n} X_i^2\right) e^{-\theta\sum_{i=1}^{n} X_i}.$$

Chesbyshev

$$P(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

(b) Note that $f(X_i) > 0$ if and only if $\theta \leq X_i \leq \theta + 1$, that is, if and only if $X_i - 1 \leq \theta \leq X_i$. Hence $L(\theta) = \prod_{i=1}^{n} f(X_i) > 0$ if and only if $\max(X_1, \ldots, X_n) - 1 \leq \theta \leq \min(X_1, \ldots, X_n)$. So we have

$$L(\theta) = \begin{cases} \left(\frac{2}{2\theta+1}\right)^n \prod_{i=1}^{n} X_i & \text{if } \max(X_1, \ldots, X_n) - 1 \leq \theta \leq \min(X_1, \ldots, X_n), \\ 0 & \text{otherwise.} \end{cases}$$

Since $L(\theta)$ is strictly decreasing in the interval $[\max(X_1, \ldots, X_n) - 1, \min(X_1, \ldots, X_n)]$ the MLE for θ is $\max(X_1, \ldots, X_n) - 1$.

Let $Y = \min(X_1, \ldots, X_n)$

$$\begin{aligned} P(Y \leq y) &= 1 - P(Y \geq y) \\ &= 1 - P(X_1 \geq y, \ldots, X_n \geq y) \\ &= 1 - \prod_{i=1}^{n} P(X_i \geq y) \\ &= 1 - (1 - F_X(y))^n \end{aligned}$$

Result: $F_Y(y) = 1 - (1 - F_X(y))^n$

$$f_Y(y) = F_Y'(y) = n\left(1 - F_X(y)\right)^{n-1} f_X(y)$$

Let $Y = \max(X_1, \ldots, X_n)$

$$\begin{aligned} P(Y \leq y) &= P(X_1 \leq y, \ldots, X_n \leq y) \\ &= \prod_{i=1}^{n} P(X_i \leq y) \\ &= F_X(y)^n \end{aligned}$$

Result: $F_Y(y) = F_X(y)^n$

$$f_Y(y) = F_Y'(y) = nF_X(y)^{n-1} f_X(y)$$

Check consistency: 2 ways. Theorem 1 or the def

$\text{Bias}(\hat{\theta}) = E_\theta[\hat{\theta}] - \theta$ $\quad SE(\hat{\theta}) = \sqrt{Var(\hat{\theta})} = \sqrt{E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2\right]}$

$$MSE(\hat{\theta}) = \text{Bias}(\hat{\theta})^2 + Var(\hat{\theta})$$

$$MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$$

(a) $\lim_{n\to\infty} Bias\left(\widehat{\theta_n}\right) = 0$

(b) $\lim_{n\to\infty} Var\left(\widehat{\theta_n}\right) = 0$

$$\lim_{n\to\infty} P\left(\left|\widehat{\theta_n} - \theta\right| > \varepsilon\right) = 0 \text{ for all } \varepsilon > 0$$

Def:

- χ_n^2 is the distribution of the sum of squares of n independent standard normal variables
- $\chi_n^2 \sim \text{Gamma}\left(\frac{n}{2}, 2\right)$
- As a consequence of this, if U and V are independent and $U \sim \chi_n^2$ and $V \sim \chi_m^2$, then $U + V \sim \chi_{n+m}^2$
- The MGF of a χ_n^2 distributed random variable is
$$M(t) = (1 - 2t)^{-n/2}$$
- Percentage points of χ_n^2 can be used to test validity of statistical models involving normal distributions

- t_k is the distribution of $\frac{C}{\sqrt{D/k}}$ where $C \sim N(0,1)$ and $D \sim \chi_k^2$ are independent
- t-distribution is similar to $N(0,1)$, but more useful than $N(0,1)$ for **small** samples
- We have $\frac{\sqrt{n}(\bar{X}-\mu)}{s} \sim t_{n-1}$ for normally distributed samples
- t-distribution can be used for hypothesis tests concerning the population mean of normally distributed samples

Let $U \sim \chi_m^2$ and $V \sim \chi_n^2$ be independent. The distribution of
$$F = \frac{U/m}{V/n}$$
is called an **F-distribution,** denoted by $F \sim F(m, n)$

- F-distribution is important for statistical tests involving variances
- A major application of the F-Distribution is Analysis of Variance (ANOVA)
- If the random variable X follows a t_n -distribution, then X^2 follows a $F(1, n)$ distribution

Solution. We first compute the CDF of X_1. Assume $\theta < x < \theta + 1$. We have

$$\begin{aligned} F_{X_1}(x) &= P(X_1 \le x) \\ &= \int_\theta^x \frac{2t}{2\theta+1} dt \\ &= \frac{2}{2\theta+1} \int_\theta^x t dt \\ &= \frac{2}{2\theta+1} \left[\frac{x^2}{2}\right]_\theta^x \\ &= \frac{x^2 - \theta^2}{2\theta+1}. \end{aligned}$$

Next, we compute the CDF of $\hat{\theta}_n = \min\{X_1, \ldots, X_n\}$. Assume $\theta < x < \theta + 1$. Using the fact that $X_1, \ldots, X_n$ are iid, we find

$$\begin{aligned} F_{\hat{\theta}_n}(x) &= P(\min\{X_1, \ldots, X_n\} \le x) \\ &= 1 - P(\min\{X_1, \ldots, X_n\} > x) \\ &= 1 - P(X_1 > x, \ldots, X_n > x) \\ &= 1 - \prod_{i=1}^n P(X_i > x) \\ &= 1 - \prod_{i=1}^n (1 - P(X_i \le x)) \\ &= 1 - \prod_{i=1}^n (1 - F_{X_i}(x)) \\ &= 1 - \left(1 - \frac{x^2 - \theta^2}{2\theta+1}\right)^n. \end{aligned}$$

Note that $\hat{\theta}_n > \theta$, since X_i [redacted] Let $\varepsilon > 0$ be given. First

$$\begin{aligned} P(|\hat{\theta}_n - \theta| > \varepsilon) &= P(\hat{\theta}_n - \theta > \varepsilon) \\ &= P(\hat{\theta}_n > \theta + \varepsilon) \\ &= 1 - F_{\hat{\theta}_n}(\theta + \varepsilon) \\ &= 1 - \left(1 - \left(1 - \frac{(\theta+\varepsilon)^2 - \theta^2}{2\theta+1}\right)^n\right) \\ &= \left(1 - \frac{(\theta+\varepsilon)^2 - \theta^2}{2\theta+1}\right)^n \\ &= \left(1 - \frac{2\varepsilon\theta + \varepsilon^2}{2\theta+1}\right)^n \end{aligned}$$

Since $0 < \varepsilon < 1$ by assumption, we have $0 < \frac{2\varepsilon\theta+\varepsilon^2}{2\theta+1} < 1$ and thus

$$1 - \frac{2\varepsilon\theta + \varepsilon^2}{2\theta+1} \in (0,1). \tag{2}$$

Using (2), we conclude

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = \lim_{n\to\infty} \left(1 - \frac{2\varepsilon\theta + \varepsilon^2}{2\theta+1}\right)^n = 0.$$

In summary, we have shown

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0 \tag{3}$$

if $0 < \varepsilon < 1$. However, if $\varepsilon \ge 1$, then $P(|\hat{\theta}_n - \theta| > \varepsilon) \le P(|\hat{\theta}_n - \theta| > \frac{1}{2})$ and thus

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0,$$

since $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \frac{1}{2}) = 0$ by (3). We conclude that

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0$$

for all $\varepsilon > 0$. Hence $\hat{\theta}_n$ is a consistent estimator for θ.