pf of 3

$$\mathbb{E}[S^2] = \mathbb{E}\Big[\underbrace{\frac{1}{N-1}\sum_{n=1}^{N}(X_n - \bar{X})^2}_{S^2}\Big]$$

trick: $\mathrm{Var}(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2$

turns out:

$$\sum_n (X_n - \bar{X})^2 = \sum_n X_n^2 - \Big(\sum_n X_n\Big)^2$$

$$= \sum_n X_n^2 - N(\bar{X})^2$$

$$= \frac{1}{N-1}\mathbb{E}\Big[\sum_n X_n^2 - N\bar{X}^2\Big]$$

$$= \frac{1}{N-1}\Big(\sum_n \mathbb{E}[X_n^2] - N\,\mathbb{E}[\bar{X}^2]\Big)$$

re-arrange short-cut: $\mathbb{E}[Z^2] = \mathrm{Var}(Z) + \mathbb{E}[Z]^2$

$$\mathbb{E}[X_n^2] = \mathrm{Var}(X_n) + \mathbb{E}[X_n]^2$$

$$E[X_n^2] = \mathrm{Var}(X_n) + E[X_n]^2$$
$$= \sigma^2 + \mu^2$$

$$E[\bar{X}^2] = \mathrm{Var}(\bar{X}) + E[\bar{X}]^2$$
$$= \sigma^2/N + \mu^2$$

$$= \frac{1}{N-1}\left(\sum_n (\sigma^2+\mu^2) - N(\sigma^2/N + \mu^2)\right)$$

$$= \frac{1}{N-1}\left(N(\sigma^2 + \cancel{\mu^2}) - N(\sigma^2/N + \cancel{\mu^2})\right)$$

$$= \frac{1}{N-1}(N-1)\sigma^2 = \sigma^2$$

Theorem: If $X_n \overset{iid}{\sim} f$ w/ MGF M

then
$$M_{\bar{X}}(t) = M(t/N)^N.$$

Pf: Recall: $M(t) = E[e^{tX}]$

Facts: $M_{aX+b}(t) = e^{tb} M_X(at)$

If $A \perp B$ then $E[AB] = E[A]E[B]$

$$M_{\bar{X}}(t) = E[e^{t\bar{X}}] = E\left[e^{t\frac{1}{N}\sum_n X_n}\right]$$

$e^a e^b = e^{a+b}$

$\prod_i e^{a_i} = e^{\sum_i a_i}$

$$= E\left[\prod_n e^{tX_n/N}\right]$$

$$= \prod_n \underbrace{E\left[e^{t/N X_n}\right]}_{M(t/N)}$$

$$= \prod_n M(t/N)$$

$$= M(t/N)^N$$

$$= M(t/N)^N$$

Ex. $X_n \overset{iid}{\sim} \text{Gamma}(\alpha, \beta)$

What's the dist of $\bar{X}$?

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x} \mathbb{1}(x>0)}{\Gamma(\alpha)}$$

$$M(t) = (1 - t/\beta)^{-\alpha}$$

$$M_{\bar{X}}(t) = M(t/N)^N$$

$$= \left[\left(1 - \frac{t}{N\beta}\right)^{-\alpha}\right]^N$$

$$= \left(1 - \frac{t}{N\beta}\right)^{-\alpha N}$$

looks like Gamma but replace

$\beta \to N\beta$

$\alpha \to N\alpha$

So MGF of $\bar{X}$ is the MGF of

Gamma$(N\alpha, N\beta)$

$\mathrm{Gamma}(N\alpha, N\beta)$

So $\bar{X}$ has this dist.

Theorem: $\bar{X}$ and S^2 for Normal Data

Let $X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

✓ (1) $\bar{X} \sim N(\mu, \sigma^2/N)$ (prove)

✓ (2) $\bar{X} \perp S^2$ (later)

✓ (3) $\frac{N-1}{\sigma^2} S^2 \sim \chi^2(N-1)$ (sketch)

chi-squared dist w/ N-1 degrees of freedom

---

pf of (1)

Use MGF theorem.

(MGF of $N(\mu, \sigma^2)$

$M(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2})$

Use MGF theorem.

$M(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$

$M_{\bar{X}}(t)$

$= M(t/N)^N$

$= \exp\left(\mu t/N + \frac{\sigma^2 (t/N)^2}{2}\right)^N$

$(e^a)^b = e^{ab}$

$= \exp\left(\mu t + \frac{\sigma^2 t^2}{2N}\right)$

$= \exp\left(\mu t + \frac{(\sigma^2/N) t^2}{2}\right)$

MGF of $N(\mu, \sigma^2/N)$

So $\bar{X} \sim N(\mu, \sigma^2/N)$

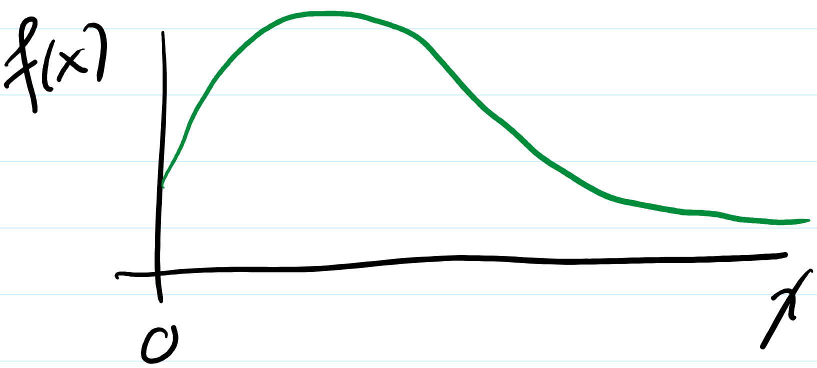
Chi-Squared dist

$\chi^2(k)$ degrees of freedom (df)

$Z \sim \chi^2(k)$ degrees of (dof)

$k > 0$

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2 - 1} e^{-x/2} \mathbb{1}(x>0)$$



$E[Z] = k$, $Var(Z) = 2k$

Facts:

1. $Z \sim N(0,1)$ then $Z^2 \sim \chi^2(1)$
2. If $Z_1 \sim N(0,1)$, $Z_2 \sim N(0,1)$, $Z_1 \perp Z_2$

   then $Z_1^2 + Z_2^2 \sim \chi^2(2)$

then

$$Z_1^2 + Z_2^2 \sim \chi^2(2)$$

Generally, if $Z_i \overset{iid}{\sim} N(0,1)$ then

$$\sum_{i=1}^{N} Z_i^2 \sim \chi^2(N)$$

3) If $Y_n \overset{indep}{\sim} \chi^2(k_n)$ then

$$\sum_{n=1}^{N} Y_n \sim \chi^2\left(\sum_n k_n\right)$$

#3 in above thm:

$$\frac{N-1}{\sigma^2} S^2 \sim \chi^2(N-1)$$

$$S^2 = \frac{1}{N-1} \sum_n (X_n - \bar{X})^2$$

so $\frac{N-1}{\sigma^2} S^2 = \sum \frac{(X_n - \bar{X})^2}{\sigma^2}$

Kinda like

So $\frac{N-1}{\sigma^2} S = \sum_n \frac{(x_n - \bar{X})^2}{\sigma^2}$

$$= \sum_n \left(\frac{x_n - \bar{X}}{\sigma}\right)^2$$

Kinda like μ

cheating

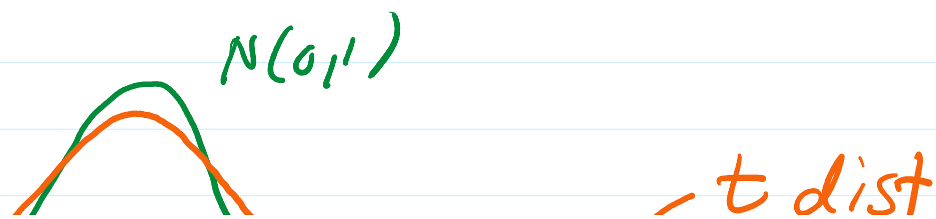
$$\approx \sum_n \underbrace{\left(\frac{x_n - \mu}{\sigma}\right)^2}_{N(0,1)^2} \sim \chi^2(N)$$

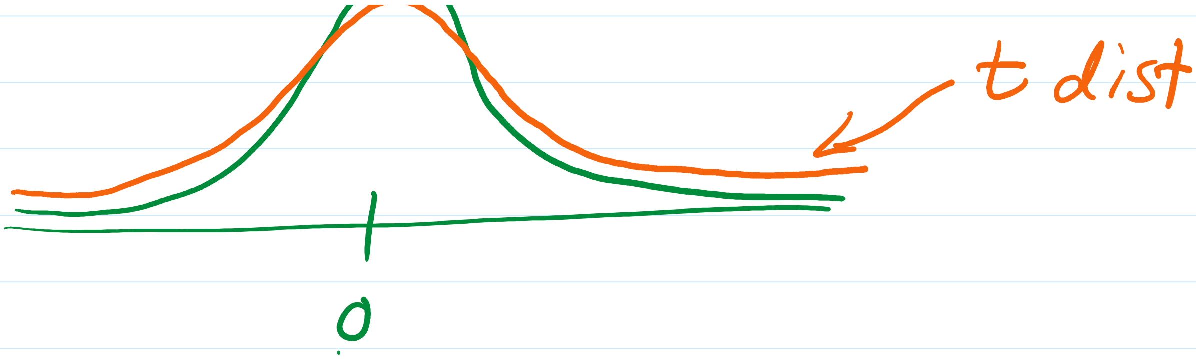
Penalty for replacing $\bar{X}$ w/ μ is we lose a dof.

So $\boxed{\frac{N-1}{\sigma^2} S^2 \sim \chi^2(N-1)}$

t-distribution

like a $N(0,1)$ but w/ fatter tails





One parameter k (called dof)

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}} \frac{1}{\left(1+x^2/k\right)^{\frac{k+1}{2}}}$$

$$\forall x \in \mathbb{R}$$

Fact: $\left.\begin{array}{l} U \sim N(0,1) \\ V \sim \chi^2(k) \end{array}\right\} U \perp V$

then $\dfrac{U}{\sqrt{V/k}} \sim t(k)$

If $X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

If $X_n \sim N(\mu, \sigma^2)$

(1) $\bar{X} \sim N(\mu, \sigma^2/N)$

(2) $\bar{X} \perp S^2$

(3) $\frac{N-1}{\sigma^2} S^2 \sim \chi^2(N-1)$

$S = \sqrt{S^2}$

So $\boxed{\frac{\bar{X}-\mu}{S/\sqrt{N}} \sim t(N-1)}$

$Z \sim N(a, b^2)$

$\frac{Z-a}{b} \sim N(0,1)$

pf $U = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}} \sim N(0,1)$

indep

$V = \frac{N-1}{\sigma^2} S^2 \sim \chi^2(N-1)$

$$\frac{U}{\sqrt{V/k}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}} \bigg/ \sqrt{\frac{\cancel{N-1}}{\sigma^2} S^2 \Big/ \cancel{N-1}}$$

$$= \frac{\dfrac{\bar{X}-\mu}{\cancel{\sigma}/\sqrt{N}}}{\sqrt{S^2}/\sigma} = \frac{\bar{X}-\mu}{(S/\sqrt{N})} \sim t(N-1)$$

Generally we'll work w/ parameterized families of dists, e.g.

- $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$
- $\mathrm{Exp}(\lambda)$, $\lambda > 0$
- $U(0, \theta)$, $\theta > 0$

Exponential Families

Assume we have a fam of dists parameterized by $\theta \in \Theta \subset \mathbb{R}$ so that

$$X_n \overset{iid}{\sim} f_\theta$$

$$X_n \overset{iid}{\sim} F_\theta$$

where

data, not θ

$$f_\theta(\underset{\sim}{x}) = h(\underset{\sim}{x})\, c(\theta) \exp\left(T(\underset{\sim}{x})\, w(\theta)\right)$$

θ not data

then we say the X_ns are from an exponential family of distr.

__Ex.__ Poisson, Exp, Normal, Gamma, ...
