BSCMA1004

STATISTICS II NOTES

WEEK 9 NOTES

IITM B.S Degree

PREPARED BY

Vehaan Handa, IIT Madras

Stats II → Week IX Notes

★ Statistical Problems

Who is the best IPL captain? You have to look at their qualities. But not easy to be quantified. What data is available? Scoresheets? Sampling of experts? Take data. Analyse. Conclude.

★ Parameter Estimation

Estimate the parameter of some distribution given some samples.

Eg. n Bernoulli(p) trials. p unknown.

One set of samples (n=10) : 1,0,0,1,0,1,1,1,0,0.
Guess p?

So we observe some iid samples from distribution, and then we are required to estimate a parameter.

$$X_1, \ldots, X_n \sim \text{iid } X$$

- X has a dist. described by some parameters $\theta_1, \theta_2 \ldots$
  - Parameters take real values, $\theta_i \in \mathbb{R}$

What is $\theta_1, \theta_2 \ldots$?

Estimation for a parameter θ:
- Function of samples $\hat{\theta}(X_1, \ldots, X_n)$
- Notation : $\hat{\theta}$ is an estimator for parameter θ

DATE

- θ is a PARAMETER (constt), not RV
- But $\hat{\theta}$ is an ESTIMATOR (fn. of n RV's), $\therefore$ it is a RV.

Eg. Bernoulli(p) trials.

$$X_1, \ldots, X_n \sim \text{iid Bernoulli}(p)$$

Parameter : p.

$\hat{p}1 = 1/2 \qquad \hat{p}2 = (X_1 + X_2)/2 \qquad \hat{p}3 = (X_1 + \ldots + X_n)/n$

Infinite no. of estimators are possible.

- Good estimators?
- How to design them?

★ Error in Estimation

θ : parameter. $\hat{\theta}(X_1, \ldots, X_n)$: estimator

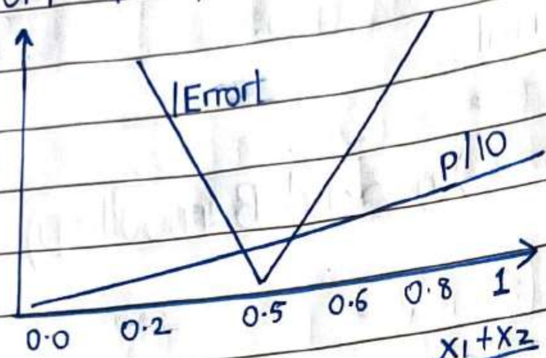
Parameter will be in a certain range. Estimator error should be low over the entire range.
Quantify 'low'? Error small compared to p.
10% or lower error : $|\text{Error}| \leq p/10$

10 samples of Ber(p) : 1, 0, 0, 1, 0, 1, 1, 1, 0, 0
$\hat{p}_1 = 0.5 \quad \hat{p}_2 = 0.5 \quad \hat{p}_3 = 0.5$
Another round : 1, 0, 0, 1, 0, 1, 0, 1, 0, 0
$\hat{p}_1 = 0.5 \quad \hat{p}_2 = 0.5 \quad \hat{p}_3 = 0.4$

$\hat{p}_3$ is good. $\hat{p}_1$ is fixed. $\hat{p}_2$ varies a lot.

$\hat{p_1} = 0.5$ • Error $= 1/2 - p$

$P(|\text{error}| > p/10) = 1$ if $p < 5/11$ or $p > 5/9$.



$\hat{p_2} = (X_1+X_2)/2$ • Error : $\frac{x_1+x_2}{2} - p$

$P(|\text{Error}| > p/10) = 1$ if $p < 5/11$ or $5/9 < p < 10/11$

X_1	X_2	$e = \frac{x_1+x_2}{2} - p$	$Pr(Err = e)$
0	0	$-p$	$(1-p)^2$
0	1	$1/2 - p$	$p(1-p)$
1	0	$1/2 - p$	$p(1-p)$
1	1	$1-p$	p^2

$\hat{p_3} = (X_1 + \dots + X_n)/n - p$

$$\boxed{P(|\text{Error} - \underset{0}{\underset{\parallel}{E[\text{Error}]}}| > \delta) \leq \frac{\overset{\frac{p(1-p)}{n}}{\text{Var(Error)}}}{\delta^2}}$$

(Chebyshev)

$$P(|Err| > p/10) \leq \frac{p(1-p)/n}{p^2/100} \leq \frac{100(1-p)}{np}$$

- For fixed p, $P \to 0$ as $n \to \infty$.
- Chebyshev : $1/n \downarrow$
- Chernoff: $\exp(n) \downarrow$
- Good design of estimator : $P(|\text{Error}| > \delta)$ will fall with
  - $E[err] \to 0$ (or $= 0$)
  - $Var[err] \to 0$ with n

Bias

$X_1, X_2, \dots, X_n \sim$ iid X, param θ

Estimator for θ : $\hat{\theta}(X_1, \dots, X_n)$ or $\hat{\theta}$.

The bias of estimator $\hat{\theta}$ for parameter θ, denoted Bias $(\hat{\theta}, \theta)$ is defined:

$$\text{Bias}(\hat{\theta}, \theta) = E[\hat{\theta} - \theta] = E[\hat{\theta}] - \theta$$

- Bias is expected value of error.
- An estimator with bias $= 0$ is said to be an unbiased estimator.

Risk (squared error)

$X_1 \dots, X_n \sim$ iid X, param θ

- Estimator for θ : $\hat{\theta}(X_1, \dots, X_n)$ or $\hat{\theta}$ in short.

Squared error risk of estimator $\hat{\theta}$ for parameter θ denoted Risk $(\hat{\theta}, \theta)$ is defined:

$$\text{Risk}(\theta, \hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

$\because$ Error $= \hat{\theta} - \theta$, risk is expected value of sq. error, also called mean square error (MSE)

Squared error risk $= 2^{nd}$ moment of error.

Variance

$Var(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])^2]$ i.e variance of estimator.

- Variance of error: $Error = \hat{\theta} - \theta$
Error is a translated version of estimator $\hat{\theta}$

$$Var(error) = Var(\hat{\theta})$$

Bias Variance Tradeoff

$X_1, \ldots, X_n \sim$ iid X, parameter θ

~~Estimator~~

$$\boxed{\text{The risk of the estimator satisfies} :\rightarrow \quad Risk(\hat{\theta}, \theta) = Bias(\hat{\theta}, \theta)^2 + Var(\hat{\theta})}$$

Eg. Bernoulli (p)

$X_1, \ldots, X_n \sim$ iid $Ber(p)$

$\hat{p}_1 = 1/2$

- $Bias = 1/2 - p$, $Var = 0$, $Risk = (1/2 - p)^2$

$\hat{p}_2 = (X_1 + X_2)/2$

$E[\hat{p}_2] = E\left[\frac{X_1 + X_2}{2}\right] = 1/2\, E[X_1] + 1/2\, E[X_2] = p.$

$\therefore$ Bias $= 0$

$Var = 1/4(\overbrace{Var(X_1)}^{p(1-p)} + Var(X_2)) = p(1-p)/2$

$Risk = p(1-p)/2$

classmate

DATE

$\hat{p}_3 = (X_1 + X_2 + \dots + X_n)/n$

- Bias $= 0$
- Variance $= \frac{1}{n^2}(Var(X_1) + \dots + Var(X_n)) = p(1-p)/n$
- Risk $= p(1-p)/n$

Eg. Let $X_1, \dots, X_n \sim$ iid Bernoulli(p). Consider estimator :→

$$\hat{p} = \frac{X_1 + \dots + X_n + \sqrt{n}/2}{n + \sqrt{n}}$$

Find bias, variance, risk of $\hat{p}$.

$$E[\hat{p}] = \frac{np + \sqrt{n}/2}{n + \sqrt{n}}$$

$$\text{Bias} = \frac{np + \sqrt{n}/2}{n + \sqrt{n}} - p = \frac{\sqrt{n}(1/2 - p)}{n + \sqrt{n}}$$

Now note that $\sqrt{n}/2$ plays no role in the variance

$$\text{Variance} = \frac{1}{(n + \sqrt{n})^2}(np(1-p))$$

Use bias variance tradeoff, Risk $= \frac{n}{4(n + \sqrt{n})^2}$

☆ Moments and Parameters

$X \sim f_X(x)$, params θ_1 and θ_2

Moments $E[X]$, $E[X^2]$ etc. can be expressed as functions of the parameters.

- Bernoulli (p)
  - $E[X] = p$
- Poisson (λ)
  - $E[X] = \lambda$
- Exponential (λ)
  - $E[X] = 1/\lambda$
- Normal (μ, σ^2)
  - $E[X] = \mu$
  - $E[X^2] = \mu^2 + \sigma^2$
- Gamma (α, β)
  - $E[X] = \alpha/\beta$
  - $E[X^2] = \alpha^2/\beta^2 + \alpha/\beta^2$

☆ Moments of Samples

$X_1, \ldots, X_n \sim \text{iid } X$

- Sample moments: $M_k(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i^k$

1^{st} sample moment : $m_1 = \frac{1}{n}(X_1 + \cdots + X_n)$

$$m_k = \frac{1}{n}(x_1^k + \ldots + x_n^k)$$

- M_k is a RV, m_k is value taken by it in one sampling instance.
  - If sampling is repeated, RV M_k takes diff. values. M_k exp. value $\rightarrow E[X^k]$.

classmate

DATE

★ Method of Moments

① Equate sample moments to expression for moments in terms of unknown parameters.

② Solve for unknowns.

One parameter θ usually needs one moment.
- Sample moment : m_1
- Distribution moment : $E[X] = f(\theta)$

Solve for θ from $f(\theta) = m_1$ in terms of m_1.
$\hat{\theta}$: Replace m_1 by M_1.

Two parameters θ_1, θ_2 usually needs 2 moments.
- Sample moments : m_1, m_2.
- Distribution moments : $f(\theta_1, \theta_2)$, $E[X^2] = g(\theta_1, \theta_2)$

Solve for ~~$g(\theta)$~~ θ_1, θ_2 from $f(\theta_1, \theta_2) = m_1$
$g(\theta_1, \theta_2) = m_2$

$\hat{\theta}_1, \hat{\theta}_2$: Replace m_1 by M_1, m_2 by M_2.

Eg. $X_1 \ldots, X_n \sim$ iid Bernoulli(p)

- $E[X] = p$
- Method of moments : $p = m_1$
- Estimator : $\hat{p} = M_1 = \frac{1}{n}(X_1 + \ldots + X_n)$

DATE

Eg. Poisson : $X_1, \ldots, X_n \sim$ iid Poisson(λ)

- $E[X] = \lambda$
- Method of moments: $\lambda = m_1$
- Estimator $\hat{\lambda} = M_1 = \frac{1}{n}(X_1 + \ldots + X_n)$

Eg. $X_1, \ldots, X_n \sim$ iid Exp(λ)

- $E[X] = 1/\lambda$
- Method of Moments: $1/\lambda = m_1$

Sol. $\lambda = \frac{1}{m_1}$

$$\hat{\lambda} = 1/M_1 = \frac{n}{X_1 + X_2 + \cdots + X_n}$$

Eg. Normal. $X_1, \ldots, X_n \sim$ iid Normal(μ, σ^2)

$$E[X] = \mu \quad E[X^2] = \mu^2 + \sigma^2$$

Method of moments :→ $\mu = m_1$
$\mu^2 + \sigma^2 = m_2$

So $\mu = m_1$ $\sigma = \sqrt{m_2 - m_1^2}$

$$\hat{\mu} = M_1 = \frac{X_1 + \ldots + X_n}{n}$$

$$\hat{\sigma} = \sqrt{\frac{X_1^2 + \ldots + X_n^2}{n} - \frac{(X_1 + \ldots + X_n)^2}{n^2}}$$

DATE

Eg. Gamma : $X_1, \ldots, X_n \sim \text{Gamma}(\alpha, \beta)$

$E[X] = \alpha/\beta \qquad E[X^2] = \alpha^2/\beta^2 + \alpha/\beta^2$

$\alpha/\beta = m_1 \qquad \alpha^2/\beta^2 + \alpha/\beta^2 = m_2$

$$\beta = \frac{m_1}{m_2 - m_1^2} \qquad \alpha = \frac{m_1^2}{m_2 - m_1^2}$$

$$\hat{\alpha} = \frac{M_1^2}{M_2 - M_1^2} \qquad \hat{\beta} = \frac{M_1}{M_2 - M_1^2}$$

Eg. ~~Bernoulli: $X_1, \ldots, X_n \sim$ iid Bernoulli(p)~~
Binomial: $X_1, \ldots, X_n \sim$ iid Binomial(N, p)

$E[X] = Np \qquad E[X^2] = N^2p^2 + Np(1-p)$

$Np = m_1$

$N^2p^2 + Np(1-p) = m_2$

$$p = \frac{m_1^2 + (m_1 - m_2)}{m_1} \qquad \hat{p}, \hat{N}$$

$$N = \frac{m_1^2}{m_1^2 + (m_1 - m_2)}$$

Estimation : $N(\mu, \sigma^2)$: 1.07, 0.91, 0.88, 1.07, 1.15, 1.02, 0.99, 0.99, 1.08, 1.08.

$\hat{\mu} = m_1 = (1.07 + \ldots + 1.08)/10 = 1.024$

$\hat{\sigma} = \sqrt{1.05482 - 1.024^2} = 0.079$

☆ Likelihood of iid samples

$X_1 \ldots X_n \sim$ iid X params: $\theta_1, \theta_2 \ldots$

$f_X(x)$ depends on $\theta_1, \theta_2 \ldots$

- $f_X(x; \theta_1, \theta_2 \ldots)$

Eg. Normal $(\mu, \sigma^2) \rightarrow f_X(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Likelihood of sampling $x_1 \ldots x_n$ i.e $L(x_1, \ldots, x_n)$

$$L(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i : \theta_1, \theta_2 \ldots)$$

Eg. Ber(p) : 1, 0, 0, 1, 0, 1, 1, 1, 0, 0.

$$L = p(1-p)(1-p)p(1-p)ppp(1-p)(1-p) = p^5(1-p)^5$$

Normal (μ, σ^2) : 1.07 ......, 1.08. (10 samples)

$$L = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{10} e^{-((1.07-\mu)^2 + \ldots + (1.08-\mu)^2)/2\sigma^2}$$

☆ Maximum Likelihood (ML) estimator

Maximum likelihood (ML) estimation:

$$\theta_1^*, \theta_2^*, \ldots = \arg\max_{\theta_1, \theta_2} \prod_{i=1}^{n} f_X(x_i; \theta_1, \theta_2 \ldots)$$

To maximise likelihood for some samples, find parameters.

When maximisation problem has a closed form solution, estimator can be expressed in terms of the samples. In many cases, this will need a numerical routine.

DATE

Eg. Bernoulli (p)

$$X_1 \ldots, X_n \sim \text{iid Bernoulli}(p)$$

- Samples : $x_1 \ldots, x_n$ • $x_i = 0$ or 1.
- Likelihood : $L(x_1 \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i)$

$f_X(x_i) = p$ if $x_i = 1$, or $f_X(x_i) = 1-p$ if $x_i = 0$.

$$L(x_1 \ldots, x_n) = p^w (1-p)^{n-w}$$

Maximise a fn :→ diff and equate to zero.
Maximising L, ↔ maximising $\log L$.

$$h(p) = \log L = w \log p + (n-w) \log(1-p)$$

$$\frac{d(h(p))}{dp} = w\, 1/p + (n-w) \frac{1}{1-p}(-1) = 0. \Rightarrow p = w/n$$

- ML estimation : $p^* = \arg\max_p p^w (1-p)^{n-w}$
  - How to find p maximising the exp?

Differentiate wrt p, equate to 0, solve for p.

$$p^* = w/n = \frac{x_1 + \ldots + x_n}{n}$$ → 1's in sample.

$$\boxed{\hat{p}_{ML} = \frac{X_1 + \ldots + X_n}{n}}$$ → Same as MOM or MME estimator.

PAGE

classmate

Eg. Poisson(λ)

$X_1, \ldots, X_n \sim$ iid Poisson (λ)

Likelihood : $L(x_1 \ldots x_n) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}$

$$L(x_1 \ldots x_n) = \frac{1}{x_1! \ldots x_n!} \underbrace{e^{-n\lambda} \lambda^{x_1 + \ldots + x_n}}_{\log(\cdot)}$$

$\lambda^* = \arg\max [(x_1 + \ldots + x_n) \log\lambda - n\lambda]$

$\lambda^* = \frac{x_1 + \ldots + x_n}{n}$ $\quad \hat{\lambda}_{ML} = \frac{X_1 + \ldots + X_n}{n}$ ← same as MME

Eg. Normal (μ, σ^2)

$X_1 \ldots X_n \sim$ iid $N(\mu, \sigma^2)$

$$L(x_1 \ldots x_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$L(x_1 \ldots x_n) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2}$$

ML estimation: $\mu^*, \sigma^* = \arg \min_{\mu, \sigma} \left[\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 + n \log\sigma\right]$

$\hat{\mu}_{ML} = \frac{X_1 + \ldots + X_n}{n}$ → Sample Mean → Same as MME

$\hat{\sigma}^2_{ML} = \frac{(X_1 - \hat{\mu}_{ML})^2 + \ldots + (X_n - \hat{\mu}_{ML})^2}{n}$ → Sample Variance

★ Finding MME and ML estimators

Eg. $X_1, \ldots, X_n \sim$ iid $\text{Exp}(\lambda)$

$$\hat{\lambda}_{MME} = \frac{n}{x_1 + \ldots + x_n}$$

$$L = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda(x_1 + \ldots + x_n)}$$

$$\lambda^* = \arg\max_{\lambda} \left[n \log\lambda - \lambda(x_1 + \ldots + x_n) \right]$$

$$\frac{n}{\lambda} - (x_1 + \ldots + x_n) = 0$$

$$\Rightarrow \lambda = \frac{n}{x_1 + x_2 + \ldots + x_n} \qquad \hat{\lambda}_{ML} = \frac{n}{x_1 + \ldots + x_n}$$

Eg. $\{1,2,3\}$ w.p. p_1, p_2, p_3.

$m_1 = \overline{X}_n \qquad m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \qquad X_1, \ldots, X_n \sim \text{iid} \overset{p_1\ p_2\ p_3}{\{1,2,3\}}$

MME

$m_1 = p_1 + 2p_2 + 3p_3$

$m_2 = p_1 + 4p_2 + 9p_3$

Let $p_3 = 1 - p_1 - p_2$

$m_1 = p_1 + p_2 + 3 - 3p_1 - 3p_2$

$= \underline{3 - 2p_1 - p_2}$

$2p_1 + p_2 = 3 - m_1$ — (1)

$m_2 = 9 - 8p_1 - 5p_2$

$8p_1 + 5p_2 = 9 - m_2$ — (2)

$2p_1 = 6 - 5m_1 + m_2$

$p_1 = 3 - (5/2)m_1 + (1/2)m_2$

$p_2 = 4m_1 - m_2 - 3$

$$\boxed{\begin{aligned} \hat{p}_{1MME} &= 3 - (5/2)M_1 + (1/2)M_2 \\ \hat{p}_{2MME} &= 4M_1 - M_2 - 3 \end{aligned}}$$

DATE

ML:

$$L = p_1^{\overset{1's}{w_1}} p_2^{\overset{2's}{w_2}} (1-p_1-p_2)^{n-w_1-w_2}$$

$$p_1^*, p_2^* = \arg\max_{p_1, p_2} [w_1 \log p_1 + w_2 \log p_2 + (n-w_1-w_2)\log(1-p_1-p_2)]$$

$$\hat{p}^*_{1\,ML} = \frac{w_1}{n} \quad \hat{p}^*_{2\,ML} = \frac{w_2}{n}$$

Eg. ~~Prob~~ Uniform $[0, \theta]$

$X \sim$ Uniform (a, b)

ML for $a \to \min\{x_1, \dots, x_n\}$

$b \to \max\{x_1, \dots, x_n\}$.

$X_1, \dots, X_n \sim$ iid Uniform $[0, \theta]$

$$f_X(x) = \begin{cases} 1/\theta & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

MME

$m_1 = \theta/2$

$\theta = 2m_1$

$\hat{\theta}_{MME} = 2M_1 = 2\overline{X} = \frac{2(X_1 + \dots + X_n)}{n}$

ML

$$L(X_1, \dots, X_n) = \begin{cases} 1/\theta^n & 0 < x_1 \dots x_n < \theta \\ 0 & \text{otherwise} \end{cases}$$

to maximise L, avoid 0.

$\theta \geq \max(x_1, \dots, x_n)$

$\arg\max_\theta \frac{1}{\theta^n}$

$\Downarrow$

least possible θ

$\hat{\theta}_{ML} = \max(X_1, \dots, X_n)$

$\hat{\theta}_{MME} = 2\overline{X}$

$\hat{\theta}_{ML} = \max(X_1 + \dots + X_n)$

Eg Uniform $\{1, 2, \dots, N\}$

$$X_1, \dots, X_n \overset{iid}{\sim} \text{Unif}\{1, \dots, N\}$$

MME

$m_1 = \frac{N+1}{2}$

$N = 2m_1 - 1$

$\hat{N}_{MME} = 2\overline{X} - 1$

ML

$\hat{N}_{ML} = \max(X_1, \dots, X_n)$

classmate

DATE

Eg. Gamma (α, β)

$$X_1, \ldots, X_n \overset{iid}{\sim} \text{Gamma}(\alpha, \beta), \quad f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

<u>MME</u>: $\hat{\alpha} = \dfrac{M_1^2}{M_2 - M_1^2}$ $\quad \hat{\beta} = \dfrac{M_2}{M_2 - M_1^2}$

<u>ML</u>: $L = \prod_{i=1}^{n} \dfrac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i} = \dfrac{\beta^{n\alpha}}{\Gamma(\alpha)^n} (x_1 \ldots x_n)^{\alpha-1} e^{-\beta(x_1 + \ldots + x_n)}$

$$\alpha^*, \beta^* = \arg\max_{\alpha, \beta} \; n\alpha \log\beta - n\log\Gamma(\alpha) + (\alpha-1)\log(x_1 \ldots x_n) - \beta \sum_{i=1}^{n} x_i$$

diff wrt β: $\dfrac{n\alpha}{\beta} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \alpha = \beta \dfrac{\sum_{i=1}^{n} x_i}{n}$ — ①

diff wrt α: $n\log\beta - \dfrac{n}{\Gamma(\alpha)} \Gamma'(\alpha) + \log(x_1 \ldots x_n) = 0.$

$$\frac{\sum_{i=1}^{n} \log x_i}{n} = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \log\beta \quad \text{— ②}$$

Eg. Binomial (N, p)

$$X_1, \ldots, X_n \sim \text{iid Binomial}(N, p)$$

<u>ML</u>: $L = \prod_{i=1}^{n} \binom{N}{x_i} p^{x_i} (1-p)^{N-x_i}$ $\quad \log L = \log\binom{N}{x_1} + \ldots + \log\binom{N}{x_n} + (x_1 + \ldots + x_n)\log p + (Nn - (x_1 + \ldots + x_n))\log(1-p)$

Diff wrt p: $Np = \dfrac{x_1 + \ldots + x_n}{n}$ $\quad = 0$

Diff wrt N: Complicated. $\quad \boxed{p = \dfrac{x}{N}}$