Statistical Inference

Lecture 07a

ANU - RSFAS

Last Updated: Mon Apr 17 23:34:22 2017

Principles of Data Reduction

- Scientists use information in a sample $X_1, \ldots, X_n$ to infer about an unknown parameter θ (could be a vector).
- The scientist usually wants to summarize a few key features of the data, which is usually done by computing statistics.
- A statistic $T(\boldsymbol{X}) = T(X_1, \ldots, X_n)$ defines a reduction of the data into a summary measure.
- A scientist may just wish to use or store $T(\boldsymbol{x})$ and will treat $\boldsymbol{x}$ and $\boldsymbol{y}$ the same if

$$T(\boldsymbol{x}) = T(\boldsymbol{y})$$

even though the samples may differ in some ways.

- While we typically no longer have need to store reduced versions of the data through statistics, the results can be useful for understanding models.

Sufficiency

Sufficiency Principle: If $T(\boldsymbol{X})$ is a sufficient statistic for θ, then any inference about θ should depend on the sample $\boldsymbol{X}$ only through $T(\boldsymbol{X})$.

Definition: A statistic $T(\boldsymbol{X})$ is **sufficient** for θ if the conditional distribution of the sample $\boldsymbol{X}$ given $T(\boldsymbol{X})$ does not depend on θ.

Sufficiency

$$\begin{aligned} P(\boldsymbol{X} = \boldsymbol{x} | T(\boldsymbol{X}) = T(\boldsymbol{x})) &= \frac{P(\boldsymbol{X} = \boldsymbol{x} \text{ and } T(\boldsymbol{X}) = T(\boldsymbol{x}))}{P(T(\boldsymbol{X}) = T(\boldsymbol{x}))} \\ &= \frac{P(\boldsymbol{X} = \boldsymbol{x})}{P(T(\boldsymbol{X}) = T(\boldsymbol{x}))} \\ &= \frac{p(\boldsymbol{x}|\theta)}{q(T(\boldsymbol{x})|\theta)} \end{aligned}$$

Note: $[\boldsymbol{X} = \boldsymbol{x}] \subset [T(\boldsymbol{X}) = T(\boldsymbol{x})]$

Sufficiency

- Eg. Let X_1, X_2, X_3 be a sample of size $n = 3$ from a Bernoulli distribution with parameter p (i.e., $P(X_i = 1) = p$).
- Consider the following two statistics:

$$T_1 = X_1 X_2 + X_3$$

$$T_2 = X_1 + X_2 + X_3$$

Sufficiency

- Suppose that $T_1 = X_1 X_2 + X_3 = 0$. This suggests one of the three possible outcomes:

$$\mathcal{X} = \{A = (0,0,0), B = (1,0,0), C = (0,1,0)\}$$

- Let's calculate the conditional distribution:

$$\begin{aligned} P(X_1 = 0, X_2 = 0, X_3 = 0 | T_1 = 0) &= \frac{P(X_1 = 0, X_2 = 0, X_3 = 0, T_1 = 0)}{P(T_1 = 0)} \\ &= \frac{P(X_1 = 0, X_2 = 0, X_3 = 0)}{P(A \text{ or } B \text{ or } C)} \\ &= \frac{(1-p)^3}{(1-p)^3 + 2p(1-p)^2} \\ &= \frac{1-p}{1+p} \end{aligned}$$

- Conditioning (i.e. knowing) the information from the statistics does not remove the parameter. So knowing the statistic is not enough. It is not sufficient.

Sufficiency

- Suppose that $T_2 = X_1 + X_2 + X_3 = 1$. This suggests one of the three possible outcomes:

$$\mathcal{X} = \{A = (1,0,0), B = (0,1,0), C = (0,0,1)\}$$

- Let's calculate the conditional distribution:
- We can then easily calculate the chance that the actual data set was $(0,1,0)$ as the conditional distribution:

$$\begin{aligned} P(X_1 = 0, X_2 = 1, X_3 = 0 | T_2 = 1) &= \frac{P(X_1 = 0, X_2 = 1, X_3 = 0, T_2 = 1)}{P(T_2 = 1)} \\ &= \frac{P(X_1 = 0, X_2 = 1, X_3 = 0)}{P(A \text{ or } B \text{ or } C)} \\ &= \frac{p(1-p)^2}{3p(1-p)^2} = \frac{1}{3} \end{aligned}$$

- Similar calculations show that for any value $T_2 = t$, the conditional distribution does not depend on p. So the statistic is sufficient.

Sufficiency

- Generally if we have:
  $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(\theta)$ and $T(\boldsymbol{X}) = X_1 + \cdots + X_n$, then:

$$
\begin{aligned}
P(\boldsymbol{X} = \boldsymbol{x} | T(\boldsymbol{X}) = t) &= \frac{p(\boldsymbol{x}|\theta)}{q(T(\boldsymbol{x}|\theta))} \\
&= \frac{\prod \theta^{x_i}(1-\theta)^{1-x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
&= \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
&= \frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}}
\end{aligned}
$$

- The conditional distribution does not depend on θ, thus $T(\boldsymbol{X})$ is sufficient.

Sufficiency

Theorem A: A neccessary and sufficient condition for $T(\boldsymbol{X})$ to be sufficient for a paramter θ is that the joint probability function factors in the form:

$$f(x_1, \ldots, x_n|\theta) = f(\boldsymbol{x}|\theta) = g(t|\theta)h(\boldsymbol{x})$$

Sufficiency

Proof (based on discrete distributions):

1. Suppose $T(\boldsymbol{X})$ is a sufficient statistic.

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= P_\theta(\boldsymbol{X}=\boldsymbol{x}) \\ &= P_\theta(\boldsymbol{X}=\boldsymbol{x} \text{ and } T(\boldsymbol{X})=T(\boldsymbol{x})) \\ &= P(\boldsymbol{X}=\boldsymbol{x}|T(\boldsymbol{X})=T(\boldsymbol{x}))P_\theta(T(\boldsymbol{X})=T(\boldsymbol{x})) \\ &= h(\boldsymbol{x})g(T(\boldsymbol{x})|\theta) = h(\boldsymbol{x})g(t|\theta) \end{aligned}$$

Sufficiency

Proof (based on discrete distributions):

2. Assume that a factorization exists. Then we can write the marginal distirbution of $T(\boldsymbol{x})$ as:

$$f_{T(\boldsymbol{x})}(t) = \sum_{\{\boldsymbol{x}: T(\boldsymbol{x})=t\}} h(\boldsymbol{x})g(t|\theta) = g(t|\theta) \sum_{\{\boldsymbol{x}: T(\boldsymbol{x})=t\}} h(\boldsymbol{x})$$

$$\begin{aligned} P(\boldsymbol{X} = \boldsymbol{x}|T(\boldsymbol{X}) = T(\boldsymbol{x})) = \frac{f(\boldsymbol{x}|\theta)}{q(T(\boldsymbol{x})|\theta)} &= \frac{h(\boldsymbol{x})g(T(\boldsymbol{x})|\theta)}{q(T(\boldsymbol{x})|\theta)} \\ &= \frac{h(\boldsymbol{x})g(T(\boldsymbol{x})|\theta)}{g(T(\boldsymbol{x})|\theta)\sum_{\{\boldsymbol{x}: T(\boldsymbol{x})=t\}} h(\boldsymbol{x})} \\ &= \frac{h(\boldsymbol{x})}{\sum_{\{\boldsymbol{x}: T(\boldsymbol{x})=t\}} h(\boldsymbol{x})} \end{aligned}$$

Sufficiency

- Example: Normally distributed data.

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$$

1. What are the sufficient statistic(s) when μ is unknown and σ^2 is known?
2. What are the sufficient statistic(s) when μ and σ^2 is uknown?
3. What are the sufficient statistic(s) when μ is known and σ^2 is unknown?

Exponential Families

A family of pdfs or pmfs is called an exponential family if it can be expressed as

$$f(x|\boldsymbol{\theta}) = S^*(x)d^*(\boldsymbol{\theta})exp\left(\sum_{i=1}^{k} c_i(\boldsymbol{\theta})T_i(x)\right)$$

or

$$\begin{aligned} f(x|\boldsymbol{\theta}) &= exp\left(\sum_{i=1}^{k} c_i(\boldsymbol{\theta})T_i(x) + log(S^*(x)) + log(d^*(\boldsymbol{\theta}))\right) \\ &= exp\left(\sum_{i=1}^{k} c_i(\boldsymbol{\theta})T_i(x) + S(x) + d(\boldsymbol{\theta})\right) \end{aligned}$$

Exponential Families

- The above is for a k-dimensional parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ and suitable choices of the functions $S(\cdot)$, $d(\cdot)$ $c_i(\cdot)$ and $T_i(\cdot)$ (for $i = 1, \ldots, k$) is termed a k-parameter exponential family.
- Note: it is important that the number of $c_i(\cdot)$ and $T_i(\cdot)$ functions is the same as the dimension of the parameter vector k.

Exponential Families

Eg: Poisson distribution.

$$X \sim \mathrm{Poisson}(\lambda), \qquad x = 0, 1, 2, 3, \ldots$$

$$\begin{aligned} f_X(x|\lambda) &= \frac{\lambda^x exp(-\lambda)}{x!} \\ &= exp\{x ln(\lambda) - \lambda - ln(x!)\} \end{aligned}$$

- The Poisson family is a one-dimensional exponential family with functions:

$$\begin{aligned} S(x) &= -ln(x!) \\ d(\lambda) &= -\lambda \\ c_1(\lambda) &= ln(\lambda) \\ T_1(x) &= x \end{aligned}$$

Exponential Families - Canonical Form

- If we define:

$$\eta = (\eta_1, \ldots, \eta_k) = c(\boldsymbol{\theta}) = \{c_1(\theta), \ldots, c_k(\theta)\}$$

then η is referred to as the canonical parameter for the exponential family and the density function can be written in the form:

$$f_X(x|\theta) = exp\left\{\sum_{i=1}^{k} \eta_i T_i(x) + B(\eta) + S(x)\right\}$$

- Note:

$$\theta = c^{-1}(\eta)$$

$$B(\eta) = d\{c^{-1}(\eta)\}$$

Exponential Families - Canonical Form

Eg: Poisson distribution.

$$X \sim \mathrm{Poisson}(\lambda), \qquad x = 0, 1, 2, 3, \ldots$$

$$\begin{aligned} f_X(x|\lambda) &= \frac{\lambda^x exp(-\lambda)}{x!} \\ &= exp\{xln(\lambda) - \lambda - ln(x!)\} \end{aligned}$$

- The canonical parameter is $\eta = ln(\lambda)$. So based on the inverse relationship we have:

$$f_X(x|\lambda) = exp\{x\eta - exp(\eta) - ln(x!)\}$$

Poisson Regression - Canonical Link Function

- In generalized linear models, one of the 'link' functions (the main one) is the canonical link function.
- The canonical link function is from the canonical form of an exponential family.
- Suppose we have data that may reasonably be considered from a Poisson distribution:

$$Y_1, \ldots, Y_n \overset{\text{indep.}}{\sim} \text{Poisson}(\lambda_i)$$

  - Now we want to relate the mean of Y_i to a linear function of covariates $(x_1, \ldots, x_k)$:

$$E[Y_i] = \lambda_i = exp(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki}) = \exp(\eta_i)$$

  - So we link the mean of the response (Y) to a linear function of the covariates (η) via the link function.

Sufficiency

Theorem: Let $X_1, \ldots, X_n$ be iid observations from a pdf or pmf $f(x|\boldsymbol{\theta})$ that belongs to an exponential family given by

$$f(x|\boldsymbol{\theta}) = S^*(x)d^*(\boldsymbol{\theta})exp\left(\sum_{i=1}^{k} c_i(\boldsymbol{\theta})T_i(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. Then

$$T(\boldsymbol{x}) = \left(\sum_{j=1}^{n} T_1(X_j), \ldots, \sum_{j=1}^{n} T_k(X_j)\right)$$

is sufficient for $\boldsymbol{\theta}$.

Proof: Tutorial question.

Minimal Sufficient

Corollary A: If $T(\boldsymbol{X})$ is sufficient for θ, the maximum likleihood estimate is a function of $T(\boldsymbol{X})$.

Proof: As we assume we have a sufficiant statistic, then we can factor the likelihood:

$$f(x_1, \ldots, x_n|\theta) = L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta) = g(t|\theta)h(\boldsymbol{x})$$

To maximize the likleihood, we only need to maximize $g(t|\theta)$.

Minimal Sufficient

Definition: A sufficient statistic $T(\boldsymbol{X})$ is called a minimal sufficient statistic if, for any other sufficient statistic $T'(\boldsymbol{X})$, $T(\boldsymbol{X})$ is a function of $T'(\boldsymbol{X})$.

- Not easy to use the definition to find a minimal sufficient statistic!

Theorem: Let $f(\boldsymbol{x}|\theta)$ be the pdf or pmf of a sample $\boldsymbol{X}$. Suppose there exists a function $T(\boldsymbol{x})$ such that, for every two sample points $\boldsymbol{x}$ and $\boldsymbol{y}$ the ratio

$$f(\boldsymbol{x}|\theta)/f(\boldsymbol{y}|\theta)$$

is constant as function of θ if and only if

$$T(\boldsymbol{x}) = T(\boldsymbol{y}).$$

Then $T(\boldsymbol{X})$ is a minimal sufficient statistic.

Minimal Sufficient

Proof: Suppose that $T(\boldsymbol{x})$ is a sufficient statistics for θ. Let $T'(\boldsymbol{x})$ be any other sufficient statistic.

- By the Factorization Theorem, there exists functions g' and h' such that
$$f(\boldsymbol{x}|\theta) = g'(T'(\boldsymbol{x})|\theta)h'(\boldsymbol{x})$$
- Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be an two sample points with $T'(\boldsymbol{x}) = T'(\boldsymbol{y})$, then
$$\begin{aligned} \frac{f(\boldsymbol{x}|\theta)}{f(\boldsymbol{y}|\theta)} &= \frac{g'(T'(\boldsymbol{x})|\theta)h'(\boldsymbol{x})}{g'(T'(\boldsymbol{y})|\theta)h'(\boldsymbol{y})} \\ &= \frac{h'(\boldsymbol{x})}{h'(\boldsymbol{y})} \end{aligned}$$
- The ratio does not depend on θ, therefore $T(\boldsymbol{x})$ is a minimal sufficient statistic for θ.

Minimal Sufficient

Example:

- Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} n(\mu, \sigma^2)$, with both μ, σ^2 unknown.
- Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two sample points.
- Let $(\bar{x}, s_{\boldsymbol{x}}^2)$ and $(\bar{y}, s_{\boldsymbol{y}}^2)$ be the sample means and sample variances for the samples $\boldsymbol{x}$ and $\boldsymbol{y}$.

$$
\begin{aligned}
\frac{f(\boldsymbol{x}|\mu, \sigma^2)}{f(\boldsymbol{y}|\mu, \sigma^2)} &= \frac{(2\pi\sigma^2)^{-n/2} exp(-[n(\bar{x}-\mu)^2 - (n-1)s_{\boldsymbol{x}}^2]/(2\sigma^2))}{(2\pi\sigma^2)^{-n/2} exp(-[n(\bar{y}-\mu)^2 - (n-1)s_{\boldsymbol{y}}^2]/(2\sigma^2))} \\
&= exp([-n(\bar{x}^2 - \bar{y}^2) + 2n\mu(\bar{x} - \bar{y}) - (n-1)(s_{\boldsymbol{x}}^2 - s_{\boldsymbol{y}}^2)]/(2\sigma^2))
\end{aligned}
$$

- This ratio will not depend on μ and σ^2 if and only if $\bar{x} = \bar{y}$ and $s_{\boldsymbol{x}}^2 = s_{\boldsymbol{y}}^2$.
- $(\bar{X}, S^2)$ are minimally sufficient for μ, σ^2.

Minimal Sufficient

- Note: Minimal sufficient statistics are not unique. Any one-to-one function of a minimal sufficient statistic is also minimal sufficient.
- In the previous example $(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ is also a set of minimal sufficient statistics for (μ, σ^2)

Minimal Sufficient

Theorem: Let $X_1, \dots, X_n$ be iid observations from a pdf or pmf $f(x|\boldsymbol{\theta})$ that belongs to an exponential family then

$$T(\boldsymbol{x}) = \left(\sum_{j=1}^{n} T_1(X_j), \dots, \sum_{j=1}^{n} T_k(X_j) \right)$$

is minimal sufficient for $\boldsymbol{\theta}$.

What do we know?

Theorem A (Rao-Blackwell):

- Let W be any unbiased estimator of $\tau(\theta)$.
- Let T be a sufficient statistic for θ.
- Define $\phi(T) = E[W|T]$.
- Then

$$E[\phi(T)] = \tau(\theta)$$

$$V[\phi(T)] \leq V[W]$$

- So if we have unbiased estimator and condition it on a sufficient statistic, our new statistic $\phi(T)$ has the same or smaller variance!!

Proof: Recall, that if X and Y are any two random variables:

$$E[X] = E[E(X|Y)]$$

$$V[X] = V[E(X|Y)] + E[V(X|Y)]$$

- Show that $\phi(T)$ is unbiased for $\tau(\theta)$:

$$\begin{aligned} E[W] &= \tau(\theta) \\ E[W] &= E[E[W|T]] = E[\phi(T)] = \tau(\theta) \end{aligned}$$

- Show that $V[\phi(T)] \leq V[W]$:

$$\begin{aligned} V[W] &= V[E(W|T)] + E[V(W|T)] \\ &= V[\phi(T)] + E[V(W|T)] \\ &\geq V[\phi(T)] \end{aligned}$$

- As $V(W|T) \geq 0$

- So the whole idea seems quite cool. We can potentially get better estimators. But the key seems to be that idea of sufficiency.
- What happens if we don't condition on a sufficient statistic?

Example: $X_1, X_2 \overset{\text{iid}}{\sim} n(\theta, 1)$. Consider the statistic $\bar{X}$:

$$E[\bar{X}] = \theta \qquad V(\bar{X}) = 1/2$$

- Now let's condition on X_1. This is not a sufficient statistic! Recall our new estimator is $\phi(X_1) = E[\bar{X}|X_1]$ (note the expectation):

$$
\begin{aligned}
\phi(X_1) &= E[\bar{X}|X_1] \\
&= \frac{1}{2}E[X_1|X_1] + \frac{1}{2}E[X_2|X_1] \\
&= \frac{1}{2}E[X_1|X_1] + \frac{1}{2}E[X_2] \\
&= \frac{1}{2}E[X_1|X_1] + \frac{1}{2}\theta
\end{aligned}
$$

- As $\phi(X_1)$ depends on an unknown parameter it is not even an estimator (statistic).
- Recall, conditioning on a sufficient statistic removes the parameter!

Theorem: If W is the best unbiased estimator of $\tau(\theta)$ (UMVUE), then W is unique.

Proof:

- Let W' be a second UMVUE.
- Set $W^* = \frac{1}{2}(W + W')$. Note: we have $E[W^*] = \tau(\theta)$

$$
\begin{aligned}
V(W^*) &= \frac{1}{4}V(W) + \frac{1}{4}V(W') + 2\frac{1}{2}\frac{1}{2}Cov(W, W') \\
&= \frac{1}{4}V(W) + \frac{1}{4}V(W') + \frac{1}{2}Cov(W, W') \\
&\leq \frac{1}{4}V(W) + \frac{1}{4}V(W') + \frac{1}{2}[V(W) \times V(W')]^{1/2}
\end{aligned}
$$

- Note: $V(W)$ has to equal $V(W')$ as the are both UMVUEs!

$$V(W^*) \leq \frac{1}{4}V(W) + \frac{1}{4}V(W) + \frac{1}{2}[V(W) \times V(W)]^{1/2} = V(W)$$

However, we can't have $V(W^*) < V(W)$! So $V(W^*) = V(W)$!

Complete Statistics

Definition: Let $f(t|\theta)$ be a family of of pdfs or pmfs for a statistic $T(\boldsymbol{x})$. The family of probability distributions is called complete if

$$E[g(T)] = \int g(t) f_T(t) dt = 0$$

for all θ implies that

$$P(g(T) = 0) = 1$$

for all θ.

Complete Statistic

Example:

- Suppose that T has a binomial (n, p) distribution, $0 < p < 1$.
- Let g be a function such that $E_\theta[g(T)] = 0$.

$$
\begin{aligned}
0 = E[g(T)] &= \sum_{t=0}^{n} g(t) \binom{n}{t} p^t (1-p)^{n-t} \\
&= (1-p)^n \sum_{t=0}^{n} g(t) \binom{n}{t} \left(\frac{p}{(1-p)} \right)^t \\
&= \sum_{t=0}^{n} g(t) \binom{n}{t} \left(\frac{p}{(1-p)} \right)^t \\
\Rightarrow \quad & 0 = \sum_{t=0}^{n} g(t) \binom{n}{t} r^t
\end{aligned}
$$

Complete Statistic

$$0 = \sum_{t=0}^{n} g(t) \binom{n}{t} r^t \quad \forall \ r$$

- The only way for this to occur is that $g(t) = 0 \ \ \forall \ t$.
- So we have:

$$P_p(g(T) = 0) = 1$$

- T is a complete statistic.

Complete Statistic

Theorem: Let $X_1, \ldots, X_n$ be iid observations from a pdf or pmf $f(x|\boldsymbol{\theta})$ that belongs to an exponential family then

$$T(\boldsymbol{x}) = \left(\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j) \right)$$

in addition to being minimal sufficient for $\boldsymbol{\theta}$ is also complete as long as the parameter space for Θ contains an open set in $\mathcal{R}^k$.

Lehman - Scheffe Theorem

Theorem: Let $X_1, \ldots, X_n$ be a random sample from a distribution with density function $f(x|\theta)$. If $T = T(\boldsymbol{X})$ is a complete and sufficient statistic, and $\phi(T)$ is an unbiased estimator of $\tau(\theta)$, then $\phi(T)$ is the unique UMVUE of $\tau(\theta)$.

Proof:

- Let U be any other unbiased estimator of $\tau(\theta)$.
- Let $U^* = E[U|T]$.
- Consider $h(T) = U^* - \phi(T)$. Recall: $\phi(T) = E[W|T]$. This means:

$$E[h(T)] = E[U^*] - E[\phi(T)] = 0, \quad \forall \quad \theta$$

- We know that T is complete. So:

$$h(T) = U^* - \phi(T) = 0 \quad \Rightarrow \quad U^* = \phi(T)$$

There is only one unbiased estimator of $\tau(\theta)$ that is a function of T!

- How to find UMVUEs? It seems we have an approach:
  1. Find or construct a sufficient and complete statistic T.
  2. Find an unbiased estimator W for $\tau(\theta)$.
  3. Compute $\phi(T) = E[W|T]$, then $\phi(T)$ is the UMVUE.
- Or:
  1. Find or construct a sufficient and complete statistic T.
  2. Find a function $g(T)$, where $E[g(T)] = \tau(\theta)$ (i.e. it is unbiased).
  3. Then $g(T)$ is the UMVUE.

Method 1

Example: Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$.

- $T = \sum_{i=1}^{n}$ is a sufficient and complete statistic for θ.
- Let's consider $W = X_1$. $E[W] = \theta$. So W is unbiased.
- Compute $\phi(T) = E[W|T]$.

Note: W is 0 or 1. $E[W] = 1P(X_1 = 1) + 0P(X_1 = 0)$.

$$
\begin{aligned}
E[W|T] &= P\left(X_1 = 1 \middle| T = t\right) \\
&= \frac{P\left(X_1 = 1,\ T = t\right)}{P\left(T = t\right)} \\
&= \frac{P\left(X_1 = 1,\ \sum_{i=1}^{n} X_i = t\right)}{P\left(\sum_{i=1}^{n} X_i = t\right)} \\
&= \frac{P\left(X_1 = 1,\ \sum_{i=2}^{n} X_i = (t-1)\right)}{P\left(\sum_{i=1}^{n} X_i = t\right)} \\
&= \frac{P\left(X_1 = 1\right) \times P\left(\sum_{i=2}^{n} X_i = (t-1)\right)}{P\left(\sum_{i=1}^{n} X_i = t\right)} \\
&= \frac{[\theta] \times [\binom{n-1}{t-1}\theta^{t-1}(1-\theta)^{(n-1)-(t-1)}]}{\binom{n}{t}\theta^{t}(1-\theta)^{n-t}} \\
&= \frac{t}{n} \Rightarrow \frac{T}{n} = \bar{X}
\end{aligned}
$$

$\bar{X}$ is the UMVUE of θ.