MVUE using Sufficient Statistics

1> MVUE for finite N can be found using CRLB iff an efficient estimator exists.

2> If efficient estimator exists for finite N, MLE will also find it. Moreover, even if efficient estimator does not exist for finite N, MLE is easy to find & is asymptotically efficient & consistent.

Q. How to find MVUE for finite N when efficient estimator does not exist?

Sufficiency principle:

Ex: $x[n] = A + w[n] \qquad n = 0, \dots, N-1$

$\hat{A} = \frac{1}{N} \sum x[n]$ is MVUE & MLE / efficient.

$E(\hat{A}) = A$, $Var(\hat{A}) = \sigma^2/N$

Q. Which data samples are sufficient for getting an estimator with this variance?

$S_1 = \{x[0], \dots, x[N-1]\} \longrightarrow |S_1| = N \leftarrow$ sufficient

$S_2 = \{x[0] + x[1], x[2], x[3], \dots, x[N-1]\}$

$\hookrightarrow |S_2| = N-1$ sufficient

$S_3 = \{\sum_{n=0}^{N-1} x[n]\} \longrightarrow |S_3| = 1 \rightarrow$ sufficient.

(Also minimal.)

Sufficiency Principle: If $T(x)$ is a sufficient statistic for θ, then any estimate of θ should depend only on $T(x)$. $\Leftrightarrow$ If for x, y, $T(x) = T(y)$ then estimate of θ should be same whether $X = x$ or $X = y$ is observed.

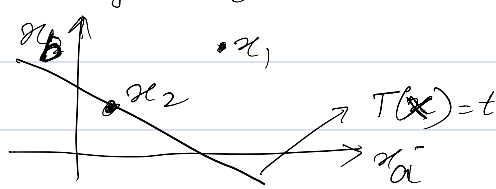
Def$\underline{\underline{n}}$: A statistic $T(X)$ is a sufficient statistic for θ if the conditional distribution of the sample X given the value of $T(X)$ does not depend on θ.

(Justification later)

Meaning: $P(X = x / T(X) = t)$ is ind of θ.

$\equiv P(X = x / T(X) = T(x))$ is ind of θ

[since if $T(x) \neq t$ then $P(X = x / T(X) = t) = 0$.]



$P(X = x_1 / T(X) = t) = 0$

But $P(X = x_2 / T(X) = t)$

$= P(X = x_2 / T(X) = T(x_2)) > 0$

[Disclaimer: All arguments about conditional prob should be done only for discrete distributions.]

For $\overset{\text{both discrete}}{\text{+ cont.}}$ distributions, we have the equivalent def$\underline{\underline{n}}$:

Def$\underline{\underline{n}}$: Let $p(x; \theta)$ be pdf/pmf of X & $q(t; \theta)$ be pdf/pmf of $T(X)$. Then

$T(X)$ is a sufficient statistic for θ, if $\forall\, x$ (in sample space), $\dfrac{p(x;\theta)}{q(T(x);\theta)}$ is independent of θ.

The equivalence of these two def^{n}s is easy to see for discrete dist (skipped for cont.)

$$P_\theta(X=x \,/\, T(X)=T(x)) = \frac{P_\theta(X=x \text{ and } T(X)=T(x))}{P_\theta(T(X)=T(x))}$$

In general, parameterized by θ. But if $T(X)$ is suff. then ind. of θ.

$$= \frac{P_\theta(X=x)}{P_\theta(T(X)=T(x))} \overset{?}{=} \frac{p(x;\theta)}{q(T(x);\theta)}$$

? For dist OK
For cont $\to$?

Q. How is principle of suff. equivalent to these definitions?

From above: $p(x;\theta) = P_\theta(X=x)$

$$= P_\theta(X=x \,/\, T(X)=T(x)) \cdot P_\theta(T(\cancel{X})=T(x))$$

not there

$$= P(X=x \,/\, T(X)=T(x)) \cdot q(T(x);\theta)$$

Let $X=x$ & $X=y$ are two outcomes: with $T(x)=T(y)$. Then consider the

MLE : $\hat{\theta}_x = \arg\max_\theta \ln p(x;\theta)$

ind of θ.

$= \arg\max_\theta \left[\ln P(X=x / T(X)=T(x)) \right]$

$+ \arg\max_\theta \ln P_\theta(T(X)=T(x))$

$= \arg\max_\theta \ln P_\theta(T(X)=T(x))$

equal

Similarly,

$\hat{\theta}_y' = \arg\max_\theta \ln P_\theta(T(X)=T(y))$

Here $\hat{\theta}_x \overset{?}{=} \hat{\theta}_y \implies$ MLE estimates of θ will be exactly equal if $P(X=x/T(X)=T(x))$ is ind. of θ.

The factorization we used above turns out to be a complete characterization of sufficiency:

<u>Neyman-Fisher Factorization Thm</u>: Let $p(x;\theta)$ be the pdf/pmf of X. A statistic $T(X)$ is a sufficient statistic for θ <u>iff</u> $\exists$ functions, $g(t;\theta)$ & $h(x)$ s.t. $\forall$ sample pts x & $\forall \theta$,

$$p(x;\theta) = g(T(x);\theta)\, h(x)$$

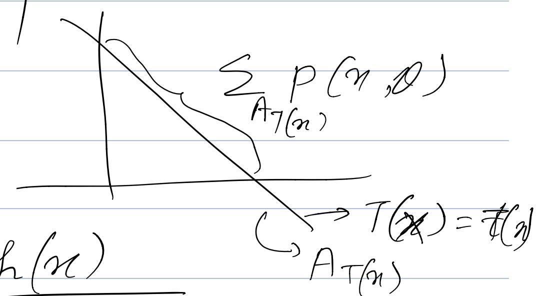
<u>Proof:</u> factorization exist $\implies$ sufficient

$$\frac{p(x;\theta)}{q(T(x);\theta)} = \frac{g(T(x);\theta)\, h(x)}{q(T(x);\theta)}$$

$$= \frac{g(T(x);\theta)\,h(x)}{\sum_{A_{T(x)}} p(x;\theta)}$$

Define

$$A_{T(x)} = \{y : T(y) = T(x)\}$$



$$= \frac{g(T(x);\theta)\,h(x)}{\sum_{A_{T(x)}} g(T(y);\theta)\,h(y)}$$

$$= \frac{g(T(x);\theta)\,h(x)}{g(T(x);\theta)\sum_{A_{T(x)}} h(y)} = \underbrace{\frac{h(x)}{\sum_{A_{T(x)}} h(y)}}_{}$$

No $\theta \Rightarrow$ sufficient.

The factorization thm is <u>very</u> useful:

Ex: $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$ (σ^2 - known) (μ - unknown)

Verify $T(X) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is sufficient statistic.

$$p(x;\mu) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{\sum (x_i - \mu)^2}{2\sigma^2}\right]$$

$$= (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{\sum (x_i - \bar{x} + \bar{x} - \mu)^2}{2\sigma^2}\right]$$

Verify Exercise.

$$= (2\pi\sigma^2)^{-n/2} \exp\left[-\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right)\cdot\frac{1}{2\sigma^2}\right]$$

Now,
$$\frac{p(x;\theta)}{q(T(x);\theta)} = \frac{(2\pi\sigma^2)^{-n/2} \exp\left[-\left(\sum (x_i - \bar{x})^2 + n(\bar{x}-\mu)^2\right)\frac{1}{2\sigma^2}\right]}{(2\pi\sigma^2/n)^{-1/2} \exp\left[-n(\bar{x}-\mu)^2/2\sigma^2\right]}$$

$$= n^{-1/2}\left(2\pi\sigma^2\right)^{-\left(\frac{n-1}{2}\right)} \exp\left(-\sum_{i=1}^{n}(x_i-\bar{x})^2/2\sigma^2\right)$$

[Recall that $T(x) = \bar{x} \sim N\left(\mu, \sigma^2/n\right)$]

Does not depend on μ. Hence $\bar{X}$ is sufficient statistic for μ.

Now use Factorization Thm:

$$p(x;\mu) = \underbrace{(2\pi\sigma^2)^{-n/2} \exp\left[-\sum(x_i-\bar{x})^2/2\sigma^2\right]}\cdot$$
$$\exp\left[-n(\underset{T(x)}{\bar{x}}-\mu)^2/2\sigma^2\right]$$

$h(x)$ does not depend on θ.

So $g(t,\mu) = \exp\left(-n(\bar{x}-\mu)^2/2\sigma^2\right)$

By factorization thm: $T(x) = \bar{x}$ is sufficient stat. for μ.

Note: $f(x;\theta) = g(T(x);\theta)\,h(x)$ for all x & for all θ

Note: The defn is useless for identifying a sufficient statistic. On the other hand sometimes it is possible to guess the factorization. & hence the sufficient stat.

In the example above, $T^1(x) = \sum x_i$ is also sufficient stat, so is $2\sum x_i$.

The data set x is always suff. (Exercise)

(they mean one-to-one & onto)

FACT: Any one-to-one function of a sufficient statistic is a sufficient statistic.

Proof: Exercise

For vector parameters, the def^n carry over easily: $T(x) = [T_1(x) \cdots T_r(x)]^T$ is said to be sufficient for $\theta = [\theta_1 \cdots \theta_p]$ (In general $r \neq p$, though usually $r = p$ in most situations), if $p(x / T(x))$ is independent of θ.

Factorization Thm: $p(x;\theta) = g(\underbrace{T(x)}_{r\times 1}, \overbrace{\theta}^{p\times 1})\, h(x)$

Ex: $x[n] = A + \underbrace{w[n]}_{N(0,\sigma^2)}$ $\quad n = 0, \ldots, N-1$

$\underbrace{A \text{ \& } \sigma^2}_{\text{unknown}}$ both unknown.

$$p(x;\theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]$$

$$= \underbrace{\frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{n=0}^{N-1}x^2[n] - 2A\sum_{n=0}^{n-1}x[n] + A^2\right)\right]}_{g(T(x);\theta)} \cdot \underbrace{1}_{1}$$

$$T(x) = \left[\sum_{n=0}^{N-1}x[n] \quad \sum_{n=0}^{N-1}x^2[n]\right]^T \rightarrow \text{suff. statistic}$$

Sufficiency for finding MVUE

Recall : $E(X) = E[E(X/Y)]$.

$Var(X) = Var[E(X/Y)] + E[Var(X/Y)]$

[Recall: $E(X/Y=y) = \int x f(x/y)\,dx$

$E(E(X/Y)) = \int E(X/Y=y) f_Y(y)\,dy$]

$Var(X/Y) = E\{[X - E(X/Y)]^2 / Y\}$

$Var(E(\cancel{X}/\cancel{Y})) = \underset{\text{w.r.t. pdf of } y}{E}\left([E(X/Y) - \underbrace{EX}_{\text{constant}}]^2\right)$

Thm: (Rao-Blackwell): Let W be any unbiased estimator of θ & let T be a sufficient statistic for θ. Define $\phi(T) = E(W/T)$.

Then (i) $E_\theta(\phi(T)) = \theta$

(ii) $Var_\theta\, \phi(T) \leq Var_\theta\, W \quad \forall \theta$

Proof: $\theta = E_\theta(W) = E_\theta[E(W/T)] = E_\theta[\phi(T)]$

i.e. if W is unbiased $\phi(T)$ is also unbiased.

$Var_\theta(W) = Var_\theta[E(W/T)] + E_\theta[Var(W/T)]$

$= Var_\theta\, \phi(T) + E_\theta[Var(W/T)]$

$\geq Var_\theta\, \phi(T) \qquad [\because Var(W/T) \geq 0]$

So $\phi(T)$ is better than W for all θ.

Q. Is $\phi(T) = E(W/T)$ an estimator? (is it a fn of X only & not of θ?)

But by defn of suff., $p(W(x)/T=t)$ is ind. of θ.

Ex: Let X_1, X_2 be iid $N(\theta, 1)$

$\bar{X} = \frac{1}{2}(X_1 + X_2)$ $E_\theta(\bar{X}) = \theta$ $Var_\theta(\bar{X}) = \frac{1}{2}$

Let $T = X_1$ (not sufficient): $\phi(X_1) = E_\theta(\bar{X}/X_1)$

Clearly, $E\phi = \theta$ & $Var(\phi) \leq Var\,\bar{X}$

But $\phi(X_1) = E_\theta(\bar{X}/X_1) = \frac{1}{2}E_\theta(X_1/X_1) + \frac{1}{2}E_\theta(X_2/X_1)$

$\underbrace{= \frac{1}{2}X_1 + \frac{1}{2}\theta}_{\text{not a valid estimator.}}$ $E(X_2/X_1) = EX_2$ (by independence)

So we should only consider estimators based on sufficient statistic ie. only $\phi(T)$'s for which $E[\phi(T)/T] = \phi(T)$ (obviously $E\phi = \theta$)

But how many such estimators are there

Thm: If W is a best (min variance) unbiased estimator of θ, then W is unique.

Proof: Let W' be another best unbiased.

Then consider $W^* = \frac{1}{2}(W + W')$. Clearly $EW^* = \theta$

& $\mathrm{Var}(W^*) = \mathrm{Var}\left(\frac{1}{2}W + \frac{1}{2}W'\right) = \frac{1}{4}\mathrm{Var}\,W + \frac{1}{4}\mathrm{Var}\,W'$
$+\frac{1}{2}\mathrm{Cov}(W, W')$

$\leq \frac{1}{4}\mathrm{Var}\,W + \frac{1}{4}\mathrm{Var}\,W' + \frac{1}{2}\left[(\mathrm{Var}\,W)(\mathrm{Var}\,W')\right]^{1/2}$
↳ Exercise

$= \mathrm{Var}\,W$ (since $\mathrm{Var}\,W = \mathrm{Var}\,W'$ by hypothesis)

Since W is best unbiased

$\mathrm{Var}_\theta W^* \geq \mathrm{Var}_\theta W$ for all θ. —— (1)

[Exercise: $\mathrm{Cov}(W, W')^2 \leq (\mathrm{Var}\,W)(\mathrm{Var}\,W')$ with equality iff $W' = \alpha W + \beta$] → Hint: Use Cauchy-Schwarz

Here, equality $\Leftrightarrow W' = \alpha(\theta)W + \beta(\theta)$

Then $\mathrm{Cov}(W, W') = \mathrm{Cov}(W, \alpha(\theta)W + b(\theta))$
$= \mathrm{Cov}(W, \alpha(\theta)W)$
$= \alpha(\theta)\,\mathrm{Var}_\theta(W)$

But $\mathrm{Cov}(W, W') = \mathrm{Var}_\theta(W)$ by (1).

So $\alpha(\theta) = 1$.

Then $\theta = EW' = E\underbrace{W}^{=\theta} + \beta(\theta) \Rightarrow \beta(\theta) = 0$.

$\Rightarrow W' = W$.

Thm: If $E_\theta W = \theta$, W is the best unbiased estimator of θ iff W is uncorrelated with all unbiased estimators of 0.

(W best $\Rightarrow$ $Cov(W, U) = 0$)

Proof: $E_\theta W = \theta$. Let U be another estimator s.t. $E_\theta U = 0 \quad \forall \theta$.

Now consider the estimator: $\phi_a = W + aU$

Clearly $E\phi_a = \theta$ <Q>: If ϕ_a better than W?

$$Var_\theta \phi_a = Var_\theta (W + aU) = Var_\theta W + 2a\, Cov_\theta(W, U) + a^2 Var_\theta U$$

Now if for some $\theta = \theta_0$, $Cov_{\theta_0}(W, U) \neq 0$, then we can choose $a \in \left(0, \mp \frac{2Cov(W,U)}{Var_{\theta_0}(U)}\right)$

Then, $Var_\theta \phi_a < Var_{\theta_0} W$

Suff: $Cov(W, U) = 0 \Rightarrow W$ best.

Let $Cov(W, U) = 0$ for all unbiased estimators of 0.

Let W' be any other estimator satisfying $E_\theta W' = E_\theta W = \theta$. $\Rightarrow E_\theta(W' - W) = 0$

Write $W' = W + (W' - W)$

$$Var_\theta(W') = Var_\theta W + Var_\theta(W' - W) + 2Cov_\theta(W, W' - W)$$
$$= Var_\theta W + Var_\theta(W' - W)$$

[Since $(W' - W)$ is an unbiased estimator of 0 & W is uncorrelated with all such estimators)

Since $Var_\theta(W' - W) \geq 0$,

$Var_\theta(W') \geq Var_\theta W$

Since W' is arbitrary, W is best unbiased est. of θ.

Q. Is this an useful characterization? It seems impossible to check!

Interpretation of Thm: An unbiased Estimator of 0 is random noise
→ If an estimator can be improved by adding random noise, then the estimator is bad.

Since the above thm is so hard to verify we <u>assume</u> that there are no unbiased estimator of zero, other than zero itself.

<u>Defn</u>: Let $f(t;\theta)$ be a family of pdfs/pmfs for a statistic $T(X)$. This set of pdfs/pmfs is called complete if $E_\theta\, g(T) = 0 \;\forall \theta$
$\Rightarrow P_\theta(g(T) = 0) = 1 \;\; \forall \theta$.

Equivalently: $T(X)$ is complete $\Rightarrow$ there is exactly one function of $T(x)$, which is an unbiased estimator.

<u>Proof</u>: Let T be complete:
Let there be two: $g(T)$ & $h(T)$ s.t.

$$Eg = Eh = \theta \quad \Rightarrow E(g-h) = 0$$
$$\Rightarrow P_\theta[g(T) - h(T) = 0] = 1 \quad \forall \theta.$$

<u>Ex</u>: $x[n] = A + w[n] \qquad w[n] \sim N(0, \sigma^2)$

We knew $T(x) = \sum x[n]$ is sufficient.

$$T \sim N(NA, N\sigma^2)$$

From def^n of completeness:

$$E_\theta(g(T)) = \int_{-\infty}^{\infty} g(T)\frac{1}{(2\pi N\sigma^2)^{1/2}} \exp\left[-\frac{1}{2N\sigma^2}(T-NA)^2\right] dT = 0 \quad \forall A$$

Let $\tau = T/N$ & $g'(\tau) = g(N\tau)$

$$\int_{-\infty}^{+\infty} g'(\tau)\frac{N}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{N}{2\sigma^2}(A-\tau)^2\right] d\tau = 0 \quad \underline{\forall A}$$

$\Rightarrow g'(\tau) = 0 \quad \forall \tau$ So $T(X)$ is complete.

(CRBLS: Rao-Blackwell-Lehmann-Scheffe)

Thm: Let T be a complete sufficient statistic for parameter θ and let $\phi(T)$ be any estimator based only on T. The $\phi(T)$ is the unique best unbiased estimator of θ.

i.e. If T is a complete suff. statistic for θ & $h(x_1, \ldots, x_n)$ is any unbiased estimator then $E(h(x_1, \ldots, x_n)/T)$ is the best unbiased estimator for θ.

<u>How to use</u>: <u>Algorithm</u>

1) Find a sufficient statistic $T(x)$ for θ using Neyman-Fisher factorization thm.
2) Determine if T is complete or not. If yes proceed, otherwise find a different T. if possible.

3) Find $\hat{\theta} = E(\bar{\theta} / T(x))$, where $\bar{\theta}$ is any unbiased estimator

OR

3') Find a function $g(T)$ that yields an unbiased estimator $\hat{\theta} = g(T)$. The MVU estimator is then $\hat{\theta}$.

Q. Why is 3' equivalent to 3? → Exercise.

Example : Incomplete Sufficient Statistic

$x[0] = A + w[0] \qquad w[0] \sim U[-\frac{1}{2}, \frac{1}{2}]$

$x[0]$ is sufficient (entire data set)

Also $x[0]$ itself is unbiased estimator of A

i.e. if $x[0]$ was complete, $E\,v(T) = 0 \Rightarrow v(T) = 0$ with Pr. 1

But let's check: assume $E\,v(T) = 0$

i.e $\int_{-\infty}^{\infty} v(T)\, p(x; A)\, dx = 0 \quad \forall A$

Here $T = x[0]$. So

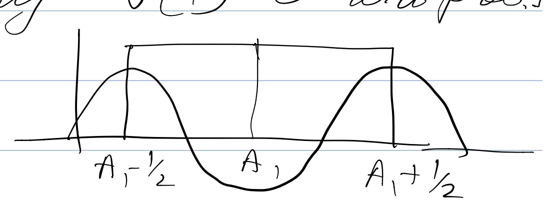
$\Leftrightarrow \int v(T)\, p(T; A)\, dT = 0 \quad \forall A$ — (1)

But $p(T; A) = \begin{cases} 1 & A - \frac{1}{2} \le T \le A + \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$

So (1) $\Leftrightarrow \int_{A-\frac{1}{2}}^{A+\frac{1}{2}} v(T)\, dT = 0 \quad \forall A$

However this does not imply $v(T) = 0$ with prob. 1.

e.g $v(T) = \sin 2\pi T$



Ex: RBLS in action : Mean of Uniform Noise

$x[n] = w[n] \qquad n = 0, 1, \cdots, N-1$

$\hookrightarrow w[n] \sim$ iid with pdf $U[0, \beta]$; $\beta > 0$.

Estimate the mean : $\theta = \frac{\beta}{2}$

Recall CRLB failed → did not satisfy regularity condition.

Try using RBLS : 1) Identify sufficient stat.

$$p(x;\beta) = \begin{cases} \frac{1}{\beta^N} & 0 < x[n] < \beta \quad n = 0, 1, \cdots, N-1 \\ 0 & \text{otherwise} \end{cases}$$

$$\Leftrightarrow \qquad = \begin{cases} \frac{1}{\beta^N} & \max x[n] < \beta, \min x[n] > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$= \underbrace{\frac{1}{\beta^N} u(\beta - \max x[n])}_{g(T(x);\beta)} \underbrace{u(\min x[n])}_{h(x)}$$

$\hookrightarrow$ unit step $u(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$

By Neyman - Fisher factorization thm, $T(x) = \max x[n]$ is a sufficient statistic.

Step 2: Completeness check:

$$\text{pdf of } T(x) : p(t;\beta) = \begin{cases} N t^{N-1} \beta^{-N} & 0 < t < \beta \\ 0 & \text{otherwise} \end{cases}$$

Exercise

Let $g(T)$ be a fn s.t. $E_\theta g(T) = 0 \quad \forall \theta$

Then, $0 = \frac{d}{d\beta} E_\beta g(T) = \frac{d}{d\beta} \int_0^\beta g(t) N t^{N-1} \beta^{-N} dt$

$$= (\beta^{-N}) \left[\frac{d}{d\beta} \int_0^{\beta} N g(t) t^{N-1} dt \right] + \left[\frac{d}{d\beta} \beta^{-N} \right] \underbrace{\int_0^{\beta} N g(t) t^{N-1} dt}_{E_\beta\, g(T) / \beta^{-N} = 0}$$

$$= \beta^{-N} N g(\beta) \beta^{N-1} + 0$$

↳ Using Liebnitz rule

$$= \underbrace{\beta^{-1}}_{\neq 0} N g(\beta) \Rightarrow \underbrace{g(\beta) = 0 \quad \forall \beta > 0}_{\text{complete stat.}}$$

(Some loss of rigour due to app. of Liebnitz rule & not showing $Pr = 1$)

So $T(x) = \max x[n]$ is complete sufficient statistic.

Step 3: Create a unbiased estimator based only on $T(x)$: Recall we had done this before: $E\{\max x[n]\} = \frac{N}{N+1}\beta$ (mean $\theta = \beta/2$)

$\Rightarrow$ For unbiasedness $g(T) = \frac{N+1}{2N} \underbrace{(\max x[n])}_{T}$

Also recall that we had calculated variance.

$$\text{Var } g(T) = \frac{1}{4N(N+2)} \beta^2$$

(We had estimated β earlier. Here we want $\theta = \beta/2$. So the extra 2.)

By RBLS $\rightarrow$ this is the unique MVUE for the mean $\theta = \beta/2$