Note 1 Basic Statistical Theory and Techniques

- introduction- prediction on future uncertainty
- Parametric model： sample-> guess distribution->calculate parameter
  - process: 获得一定量的样本数据之后，通过画图我们估计这个样本的分布，然后利用计算机计算参数

    > step 1: 获得stock data and calculate the daily return
    > step 2: visualize the distribution using histogram, found it is bell-like
    > step 3: determine the parameters

  - population->sample data(through random sampling)->estimation(investigation and estimation)->parameter(prediction)

    > for example: \theta_1=(x_1+x_2+...+x_n)/n(by tuition)同时重复这个过程，可以得到很多theta，theta_n 是随机变量
    > but why? whether theta is a good estimator?

  - How to measure the accuracy of the estimator?
    - 1.unbiased:
      - 求参数的期望

        $$\mathbb{E}[\hat{\theta}_n \mid \theta] = \theta \quad \Rightarrow \text{unbiased}$$

        otherwise, $$Bia_{\theta}(\hat{\theta}_n) = \mathbb{E}[\hat{\theta}_n \mid \theta] - \theta$$

    - 2.consistent: 注意unbiased 和 consistent 不存在等价或包含关系
      - 和概率相关

        $P(\hat{\theta}_n$ is close to $\theta)$ is close to 1.

        but how to define "close to"?

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

        注意 unbiased 和 consistent 不存在等价或包含关系.

        ① unbiased but not consistent

        $X_1 \dots X_n$, 令 $\hat{\theta}_n = X_1$

        $\therefore \mathbb{E}[\hat{\theta}_n] = \mathbb{E}[X_1] = \mathbb{E}[X]$ $\quad \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

        but $\underset{n\to\infty}{\lim} P(|X_1 - \theta| > \varepsilon) \neq 0$ $\quad \mathbb{E}[X_i^2] = Var(X_i) + (\mathbb{E}[X_i])^2 = \sigma^2 + \mu^2$

        ② consistent but not unbiased $\quad X_i \sim N(\mu, \sigma^2)$

        eg 用样本方差估计总体方法, sample: $X_1 \dots X_n \rightarrow \bar{X} = \frac{X_1 + \dots + X_n}{n}$

        $X_1 \dots X_n \sim N(\mu, \sigma^2)$ 样本方差 $s^2 = \frac{\sum (X_i - \bar{X})^2}{n}$ 总体方差 σ^2.

$$E[s^2] = E\left[\frac{\sum(x_i-\bar{x})^2}{n}\right] = \frac{1}{n}E\left[\sum x_i^2 - 2\underbrace{\sum x_i}_{n\bar{x}}\bar{x} + \sum\bar{x}^2\right]$$

$$= \frac{1}{n}\left(\sum E[x_i^2] - 2E[n\bar{x}\cdot\bar{x}] + \sum E[\bar{x}^2]\right)$$

$x_i \sim N(\mu, \sigma^2)$

$E[x_i^2] = Var(x_i) + (E[x_i])^2 = \mu^2 + \sigma^2$

$$= \frac{1}{n}\left(n(\mu^2+\sigma^2) - n\cdot E[\bar{x}^2]\right)$$

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

不好判断谁大谁小 $= \frac{1}{n}(n-1)\sigma^2 = \frac{n-1}{n}\sigma^2$. unbiased.

$\lim\limits_{n\to\infty} P(|s^2-\sigma^2| > \varepsilon)$

换个 uniform 的例子 $[0,\theta]$ $\hat{\theta}_n = \max(\theta_1 \dots \theta_n)$

$E[\hat{\theta}_n] = \underbrace{E[\max(\theta_1 \dots \theta_n)]}$ 不好计算

$= \int \hat{\theta}_n f_\theta(\hat{\theta}_n)\, d\theta$

$\therefore$ 由 F. $F(\hat{\theta}_n < y) = F(\theta_1<y, \theta_2<y \dots \theta_n<y) = F(\theta_1<y)F(\theta_2<y)\dots F(\theta_n<y)$

$= \left(\frac{y}{\theta}\right)^n$

$$f_{\hat{\theta}_n}(y) = \left[\left(\frac{y}{\theta}\right)^n\right]' = \frac{ny^{n-1}}{\theta^n}$$

$$\therefore E[\hat{\theta}_n] = \int_0^\theta y\frac{n\cdot y^{n-1}}{\theta^n}dy$$

$$= \frac{n}{\theta^n}\cdot\frac{1}{n+1}y^{n+1}\Big|_0^\theta$$

$$= \frac{n}{\theta^n}\cdot\frac{1}{n+1}\cdot\theta^{n+1} = \frac{n}{n+1}\theta. \quad \boxed{\text{biased}}$$

unconsistent $\lim\limits_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = \lim\limits_{n\to\infty} P(\theta - \hat{\theta}_n > \varepsilon) = \lim\limits_{n\to\infty} P(\hat{\theta}_n < \theta - \varepsilon)$

$\hat{\theta}_n = \max(y_1 \dots y_n) < \theta$.

$= \lim\limits_{n\to\infty} F(\theta-\varepsilon)$

$= \lim\limits_{n\to\infty} \left(\frac{\theta-\varepsilon}{\theta}\right)^n$

$= 0$ [证毕]

- 3.Mean square error and UMVUE
  - 之前都是用绝对值，但绝对值一般来说很难算，于是我们采用平方来cancel out绝对值的影响
  - UMVUE: uniformly minimum variance unbiased estimator一致最小方差无偏估计

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

$$= E[\hat{\theta}_n^2 - 2\theta\hat{\theta}_n + \theta^2]$$

$$= E[\hat{\theta}_n^2] - 2\theta\cdot E[\hat{\theta}_n] + \theta^2$$

$$= \underbrace{E[\hat{\theta}_n^2] - (E[\hat{\theta}_n])^2} + (E[\hat{\theta}_n])^2 - 2\theta\cdot E[\hat{\theta}_n] + \theta^2$$

$$= Var(\hat{\theta}_n) + [E[\hat{\theta}_n] - \theta]^2$$

$$= Var(\hat{\theta}_n) + (Bias_\theta(\hat{\theta}_n))^2$$

如果 $E[\hat{\theta}_n] = \theta$. 即 $\hat{\theta}_n$是 unbiased. $\Rightarrow$ estimator with minimum variance.

  - 多个theta可以用MSE选择better的theta
- Remark:
  - asymptotically unbiased $\lim E[\theta_n|\theta] = 0$
  - can modify biased estimator to unbiased, 样本方差❌n/(n-1)

- Interval estimation : unknown true theta, can't choose the best θ_n sample -> parameter
  - If we don't know the true value of theta? we can't find an estimator which we are sure that it is unbiased or consistent or uses the MSE
  - how about the range?If we can get the possible range, it will help a lot.
  - procedure:

    we choose $\hat{\theta_L}$ and $\hat{\theta_u} \Rightarrow P(\hat{\theta_L} \le \theta \le \hat{\theta_u}) = P$

    $\therefore [\theta_L, \theta_u] \Rightarrow (100p)\%$ confidence interval

    Example: we know population variance $= \sigma^2 \Rightarrow$ we want to estimate μ.

    Let $X_1, X_2 \dots X_n$ be the sample, $\bar{X} = \frac{X_1+X_2+\dots+X_n}{n}$ $\underline{\underline{P}}$

    由于 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

    $\therefore \frac{X_1+\dots+X_n}{n} \sim N(\mu, \frac{\sigma^2}{n})$

    $\therefore \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$. 由 $P(z_L \le \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \le z_u) = P$

    可得 U_L 和 U_u. $\Rightarrow P(\underbrace{z_L\sqrt{\frac{\sigma^2}{n}}+\mu}_{U_L} \le X \le \underbrace{z_u\sqrt{\frac{\sigma^2}{n}}+\mu}_{U_u}) = P$

    step 1: obtain a point estimator for theta, denoted by
    step 2: Determine the distribution for
    step 3: Find L and U s.t. ,在正态分布时， f是标准化的过程
    step 4: 解出L and U , 然后反解出和
  - Example注意复习！
- Parameter estimation in parameter model sample + true distribution -> parameter
  - Supposed X is assumed to follow a certain distribution with unknown theta, in most of the time, it's difficult to guess theta's formula
  - Method 1 -> Methods of moments
    - 因为已知X的分布，我们可以得到theta的公式，例如如果已知X服从poisson分布，则E[x^2]=lamda^2+lamda，其中E[X^2]可以利用样本算出来，就可以解出lamda

      $$f(\theta) = E[X^k] = \frac{X_1^k + \dots + X_n^k}{k}$$

    - Limitations:
      - may be inconsistent, such as uniformly distribution
  - Method 2 -> Maximum likelihood method
    - 因为我们已知distribution, 那么sample来自正确的参数的模型的联合概率应该是最大的
    - Likelihood function

- $L(\theta_1, \theta_2, ..., \theta_k) = \prod P(X_i = x_i|\theta_1, \theta_2, ..., \theta_k)$
- 将它看成关于theta的函数，选择function最大的theta
- quality
  - asymptotically unbiased
  - consistent
  - distribution

    n→∞时 $\hat{\theta}_n \sim N(\mu, \frac{1}{I(\theta)})$

    $I(\theta)$ 是 Fisher's information、$I(\theta) = -nE[\frac{\partial^2}{\partial \theta^2} \ln f_X(X|\theta)]$

  - n-> infinity, MLE has the smallest MSE among all the unbiased estimators

    $Var(\hat{\theta}) \geq \frac{1}{I(\theta)}$ $\hat{\theta}$ 是任意一个 unbiased estimator

- Proof of the four qualities

- Non-parametric model and empirical distribution and empirical distribution(判断sample data 是否是从这个分布中随机抽取的样本)
  - Empirical distribution
    - to check the proposed model is correct in the sense that the corresponding distribution is consistent to the distribution of sample data
    - proof

      $$F_n(x) = \frac{\text{Number of data} \leq x}{n} = \frac{\sum_{i=1}^{n} 1\{X_i \leq x\}}{n}$$

      $$1\{A\} = \begin{cases} 1 & \text{if } A \text{ happens} \\ 0 & \text{if } A \text{ doesn't happen} \end{cases} \Rightarrow \text{indicator function}$$

    - 如果样本数量足够大，那么经验分布应该趋近于实际分布

      $$F_n(x) = \frac{\sum_{i=1}^{n} 1\{X_i \leq x\}}{n} \xrightarrow{n \to \infty} E[1\{X_i \leq x\}] = P(X_i \leq x) = F_X(x)$$

      $$= \frac{\sum_{i=1}^{n} 1\{X_i \leq x\}}{n}$$

  - but how to judge the empirical distribution is close to the true distribution? Percentile and QQ-plot
  - percentile
    - definition

我们定义 π_p 是 $(100p)\%$ 分位点，$\Leftarrow F_X(\pi_p) = P(X \le \pi_p) = p$

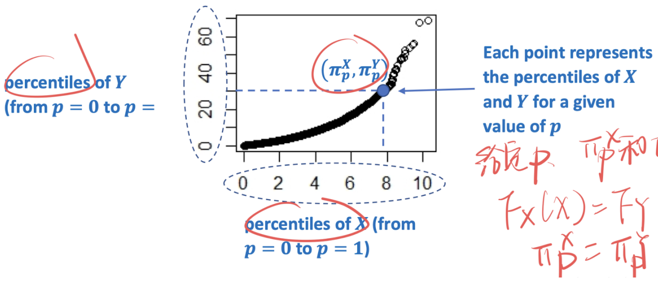
$$\begin{cases} X \text{ is continuous r.v.} \Rightarrow \text{have unique solution} \\ X \text{ is discrete r.v} \Rightarrow \text{may not have solution.} \end{cases}$$

$$\Rightarrow \pi_p = \sup\{x \mid F_X(x) < p\}$$

$$\text{or } \pi_p = \inf\{x \mid F_X(x) \ge p\}$$

- If X 和 Y具有相同的分布函数，那他们的分位点应该是相同的，将两个分布的分布点画在一张图内，能得到一条直线

$$\left\{\begin{matrix} X \text{ and } Y \text{ has same} \\ distribution\ F_X(z) = F_Y(z) \end{matrix}\right\} \Leftrightarrow \left\{\begin{matrix} \pi_p^X = \pi_p^Y \\ for\ all\ p \in (0,1) \end{matrix}\right\}.$$



给定p π_p^X和π_p^Y的值

$F_X(x) = F_Y(x)$

$\pi_p^X = \pi_p^Y$

$$F_X(\pi_p^X) = p \Rightarrow P(X \le \pi_p^X) = P\left(\frac{X-\mu}{\sigma} \le \frac{\pi_p^X - \mu}{\sigma}\right)$$

$$= P\left(Z \le \frac{\pi_p^X - \mu}{\sigma}\right) = p.$$

$$\therefore \frac{\pi_p^X - \mu}{\sigma} = \pi_p^Z$$

$$\therefore \pi_p^X = \pi_p^Z \sigma + \mu$$

- Hypothesis Testing on normality check
    - we don't have the explicit definition on "close" because of QQ plot
    - test statistic
        - 用来判断F_n(x)和F_x(x)之间的difference
    - KS-test
        - H0: The sample data is drawn from the probability distribution ; H1: The sample data is NOT drawn from the probability distribution
        - D*是自己设置的阈值，如果D_n>D,证明difference 太大了，拒绝原假设即拒绝H_0,接受H_1
        - 这里添加了一些微小的噪声，为了保证结果的鲁棒性

```
> x<-x+0.0000000000000001*rnorm(length(x),0,1)
```

        - t分布

$$T_v = \frac{Z}{\sqrt{\frac{X_v^2}{v}}} \quad \begin{matrix} \sim Z(0,1) \\ \sim \chi^2 \end{matrix}$$

$\therefore T_v \sim$ student. t分布.

but it is 对称的. degree of freedom is v.

$\Rightarrow$ general t distribution.

$$X = \mu + \lambda T_v \qquad \mathbb{E}[X] = \mu. \qquad Var(X) = \frac{n}{n-2}\lambda^2$$

其中, μ称为 location parameter. λ为 scale parameter

v可以 indicate tail $\Rightarrow$ $v\downarrow$ tail heavier

- Anderson-Darling Test
  - 很多时候我们非常关注尾部的拟合，这时再使用KS-test就不再合适了，因为尾部的概率很小，我们没办法做出判断
  - 于是，我们在期望中添加一个权重函数，使得我们可以put more weights to the tails
  - 同时，我们不再使用绝对值了，为了消除绝对值的影响，我们采用平方
  - H0: The sample data is drawn from the probability distribution ; H1: The sample data is NOT drawn from the probability distribution

$$D_n = \mathbb{E}[(F_n(x) - F_X(x))^2 \phi(x)]$$

$$= \int_{-\infty}^{+\infty} [F_n(x) - F_X(x)]^2 \cdot \phi(x) \cdot f_X(x)\,dx$$

我们看下列随机变量 $\{\text{number of data st. } X_i \le x\} = \sum_{i=1}^{n} 1\{X_i \le x\}$

服从 binomial distribution. $F_X(x) = P(X \le x) = p$ (二项分布)

$$F_n(x) = \frac{\{\text{number of data st. } X_i \le x\}}{n} \text{ (经验分布)}$$

$$\mathbb{E}[F_n(x)] = \mathbb{E}\left[\frac{\{\text{number of } X_i \le x\}}{n}\right] = \mathbb{E}\left[\frac{\sum_{i=1}^{n} 1\{X_i \le x\}}{n}\right]$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n} E[1\{X_i \le x\}]$$

$$= \frac{1}{n}\cdot\sum_{i=1}^{n} p = p = F_X(x)$$

$$\therefore \mathbb{E}[(F_n(x) - F_X(x))^2] = Var(F_n(x)) + \underbrace{(\mathbb{E}[F_n(x) - F_X(x)])^2}_{=0}$$

$$= Var(F_n(x))$$

$$\mathbb{E}[(F_n(x) - F_X(x))^2] = Var(F_n(x))$$

$$= Var\left(\frac{\{\text{number of } X_i \le x\}}{n}\right)$$

$$= \frac{1}{n^2} Var\{\text{number of } X_i \le x\}$$

$$= \frac{1}{n^2} Var\left\{\sum_{i=1}^{n} 1\{X_i \le x\}\right\}$$

$$= \frac{1}{n^2}\cdot\sum_{i=1}^{n} Var\{1\{X_i \le x\}\}$$

$$= \frac{1}{n^2}\cdot\sum_{i=1}^{n} p(1-p)$$

$$= \frac{1}{n} p(1-p)$$

$$= \frac{F_X(x)(1-F_X(x))}{n}$$

0	1
$X_i > x$	
$1-p$	p

$\therefore E = p$.

$$\therefore Var = p^2(1-p) + (1-p)^2 \cdot p$$

$$= p(1-p)(p + 1-p)$$

$$= p(1-p)$$

$$\therefore \text{我们令 } \phi(x) = \frac{n}{F_X(x)(1-F_X(x))}$$

$$\mathbb{E}[(F_n(x)-F_X(x))^2] = \mathbb{E}\left[\frac{F_X(x)(1-F_X(x))}{n}\right]$$

两边在期望中同时乘以 $\phi(x) = \frac{n}{F_X(x)(1-F_X(x))}$

$$\mathbb{E}[(F_n(x)-F_X(x))^2\phi(x)] = \mathbb{E}[1] = 1$$

同时 $\phi(x) = \frac{n}{u(1-u)} = \frac{n}{-u^2+u}$

在 $u=0/1$ 时 $-u^2+u$ min, $\phi(x)$ max.

∴达到了将更大的权重赋给tail的目的

$$\therefore D_n^* = n\int_{-\infty}^{+\infty} \frac{(F_n(x)-F_X(x))^2}{F_X(x)(1-F_X(x))} f_X(x)dx$$

但显然这个公式过于复杂，不能用来计算

- How to work with Anderson-Darling Test
    - 利用分布函数服从于均匀分布

我们考虑 $Y = F_X(x)$

$\therefore F_Y(u) = P(Y\le u) = P(F_X(x)\le u) = P(X\le F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u$

$\therefore F_Y(u)$ 是 uniform distribution. $u\in[0,1]$ $Y\sim U[0,1]$

此时 $F_n(u) = \frac{k}{n}$ $F_u(u) = u$

$$\therefore D_n^* = \int_{-\infty}^{+\infty} \frac{(F_n(u)-F_u(u))^2}{F_u(u)(1-F_u(u))} f_u(u)du = \int_{-\infty}^{+\infty}\frac{(\frac{k}{n}-u)^2}{u(1-u)}\cdot 1\,du$$

$$= n\sum_{k=0}^{n}\left(\int_{u_k}^{u_{k+1}}\frac{(\frac{k}{n}-u)^2}{u(1-u)}du\right)$$

$$= n\sum_{k=0}^{n}\left(\int_{u_k}^{u_{k+1}}\left(\frac{\frac{k^2}{n^2}-2u\cdot\frac{k}{n}+u^2}{u(1-u)}du\right)\right)$$

$$= n\sum_{k=0}^{n}\left(\int_{u_k}^{u_{k+1}}\left(\frac{k^2}{n^2u(1-u)}-\frac{2\frac{k}{n}}{1-u}+\frac{u}{1-u}du\right)\right) \quad -1+\frac{1}{1-u}$$

$u_0=0$ $u_{n+1}=1$

$$= n\sum_{k=0}^{n}\left(\int_{u_k}^{u_{k+1}}\frac{k^2}{n^2}\left(\frac{1}{u}+\frac{1}{1-u}\right)+\frac{-2\frac{k}{n}}{1-u}+\frac{u}{1-u}du\right)$$

$$= n\sum_{k=0}^{n}\int_{u_k}^{u_{k+1}}\left(\frac{(\frac{k}{n})^2}{u}+\frac{(\frac{k}{n})^2}{1-u}-\frac{(\frac{2k}{n}-1)}{1-u}-1\right)du$$

$$= n\sum_{k=0}^{n}\left[(\tfrac{k}{n})^2(\ln u_{k+1}-\ln u_k)+(\tfrac{k}{n}-1)^2(-\ln(1-u_{k+1})+\ln(1-u_k))-(u_{k+1}-u_k)\right]$$

$$= n\sum_{k=0}^{n}\Big[\frac{k^2}{n^2}\ln u_{k+1}-\frac{k^2}{n^2}\ln u_k+\frac{k^2}{n^2}\ln(1-u_k)-\frac{k^2}{n^2}\ln(1-u_{k+1})$$
$$-\frac{2k}{n}\ln(1-u_k)+\frac{2k}{n}\ln(1-u_{k+1})+\ln(1-u_k)-\ln(1-u_{k+1})$$
$$\underbrace{-u_{k+1}+u_k}\Big]$$

?

$-u_{n+1}+u_n-u_n+u_{n-1}\cdots\cdots-u_1+u_0$

$=-1$

这里让每个区间 (u_k, u_{k+1}) 内只有1个点

$$= -\sum_{k=1}^{n}\frac{2k-1}{n}[\ln u_k+\ln(1-u_{n+1-k})]-n$$

$$= -\sum_{k=1}^{n}\frac{2k-1}{n}[\ln F_X(X_k)+\ln(1-F_X(X_{n+1-k}))]-n$$

- Monte Carlo Simulation -> calculate E[f(x)] <-提前知道真实的密度函数
    - Process
        - Step 1: generate a set of random number
            - If normal, lognormal, exponential -> rnorm, rlnorm, rexp
            - if not,

- Step 1: generate a random number U from uniform distribution
- Step 2: $F_X(Y) = U$, 可以得到Y

- Step 2: calculate $f(X_1), f(X_2), ..., f(X_n)$
- Step 3: $E[f(X)] = \frac{f(X_1)+f(X_2)+...+f(X_n)}{n}$

- 我们很难知道真实的value，也就没办法知道我们的估计误差，我们选择定义

cancel out 0/or measure.

$$SE = \sqrt{\mathbb{E}[(\bar{X} - \mathbb{E}[\bar{X}])^2]} = \sqrt{Var(\bar{X})} = \sqrt{\frac{1}{n}Var(X)} \approx \sqrt{\frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)}$$

$$= \sqrt{Var\left(\frac{X_1+...+X_n}{n}\right)}$$

$$= \sqrt{\frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n^2(n-1)}}.$$

$$= \sqrt{\frac{1}{n^2} Var(X_1+...+X_n)}$$

$$= \sqrt{\frac{1}{n} Var(X)}$$

- Drawbacks:
  - 如果分布非常复杂，我们利用抽取uniform distribution，但是这个用到了反函数，这个反函数有的时候我们是没办法计算的
  - 如果想要得到准确的分布，足够大的n才能减小误差，但这需要很大的计算量
  - 以下几种方法可以解决这些问题

- Rejection Sampling
  - $E[h(x)]$,$h(x)$非常复杂，如何换一个比较简单的密度函数$g(x)$来抽取样本
  - $f_{x,y}(x,y)$服从于uniform distribution，定义域是整个密度函数的面积
  - Process
    - Step 1: $Mg_X(x) \geq f_X(x)$
    - Step 2: 先抽取$X^* < -g_X(x)$，然后抽取$Y^* < -U[0, Mg_X(X^*)]$
    - Step 3: 如果$Y^* > f_X(X^*)$, reject; 否则，我们则选择抽取这个sample X^*
  - Proof:

$$F_{X^*}(x) = P(X^* \leq x \mid Y^* \leq f_X(X^*))$$

$$= \frac{P(X^* \leq x, Y^* \leq f_X(X^*))}{P(Y^* \leq f_X(X^*))}$$

$$= \frac{\int_0^x P(Y^* \leq f_X(z) \mid X^* = z)\,dz}{\int_{-\infty}^{+\infty} P(Y^* \leq f_X(z) \mid X^* = z)\,dz}$$

$$= \frac{\int_{-\infty}^x g_X(z) \cdot P(Y \leq f_X(X^*) \mid X^* = z)\,dz}{\int_{-\infty}^{+\infty} g_X(z) \cdot P(Y \leq f_X(X^*) \mid X^* = z)\,dz} \qquad = \frac{f_X(z)}{Mg_X(z)}$$

$$= \frac{M\int_{-\infty}^x f_X(z)\,dz}{M\int_{-\infty}^{+\infty} f_X(z)\,dz_{=1}} = \frac{F_X(x)}{1} = F_X(x)$$

  - Remark:

- $g_X(x)$必须和$f_X(x)$同正同负
- M如果太大，会导致抽取的uniform sample没那么有效，会拒绝大量的sample，所以选择一个有效的M，才最合适

  $$M = \sup_{x\in(-\infty,\infty)} \frac{f_X(x)}{g_X(x)}.$$

- 同时，拒绝域也可以修改，比如
  - 1: uniform 抽取规则修改
    - Step 1: $Mg_X(x) \geq f_X(x)$
    - Step 3: 如果$Y^* > \frac{f_X(X^*)}{Mg_X(X^*)}$, reject; 否则，我们则选择抽取这个sample X^*
    - Step 2: 先抽取$X^* < -g_X(x)$，然后抽取$Y^* < -U[0,1]$
  - 2: 阈值修改
    - 有时题目中给$f_X(x) = Cf_X^*(x)$, C是为了将$f_X(x)$积分=1，其实不必一定要求出C，将剩下步骤的$f_X(x)$都改成$f_X^*(x)$也可以，详情看example 23

- Importance Sampling
  - 举例子来看，为什么X变大，我们的error会变大呢？
    - 因为X是我们规定的strike price, 现在的价格是100，如果X=500，这种情况是非常不常见的，也就导致sample会特别少，sample 很少就会影响我们的估计精度，怎么解决这个问题呢？

      $$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t},$$

    - 将不太常见的sample赋予一个较大的被抽中的概率，在这个题目中，将WT向右平移，使得他中心从0变成某个值，从而给不常见的X=ST赋予一个较大的权重，我们就将X=ST=WT的mu， 从而可以得出mu等于某个数值

      $$X = S_0 e^{\left(r - \frac{\sigma^2}{2}\right)T + \sigma\mu} \Rightarrow \mu = \frac{\ln\left(\frac{X}{S_0}\right) - \left(r - \frac{\sigma^2}{2}\right)T}{\sigma}.$$

- Antithetic Variate
  - 已有一个sample，如何通过这个sample获得另一个
  - $X_i \to F(X_i) = U_i \to 1 - U_i = F(Y_i) \to Y_i$
  - 证明X_i和Y_i来自同一分布

$$
\begin{aligned}
F_{Y_i}(y) &= P(Y_i \le y) \\
&= P(F_X(Y_i) \le F_X(y)) \\
&= P(1-U_i \le F_X(y)) \\
&= P(U_i \ge 1-F_X(y)) \\
&= 1-P(U_i < 1-F_X(y)) \qquad U_i \sim U[0,1] \\
&= 1-(1-F_X(y)) \\
&= F_X(y)
\end{aligned}
$$

- 为什么这样就能减小估计误差?
    - 直观理解：这样变成了2n个数据，当然可以减小误差
    - 数学计算：

之前 $\mathbb{E}[h(X)] = \frac{h(x_1)+\dots+h(x_n)}{n}$ $\qquad SE_{old} = \sqrt{\frac{Var(h(X))}{n}}$

$\mathbb{E}[\hat{h}(X)] = \frac{\hat{h}_1+\hat{h}_2+\dots+\hat{h}_n}{n} = \frac{\frac{h(x_1)+h(y_1)}{2}+\frac{h(x_2)+h(y_2)}{2}+\dots+\frac{h(x_n)+h(y_n)}{2}}{n}$

$\mathbb{E}\left[\frac{h(x_i)+h(y_i)}{2}\right] = \frac{1}{2}\left(\mathbb{E}[h(x_i)]+\mathbb{E}[h(y_i)]\right) = \mathbb{E}[h(X)]$

$SE_{new} = \sqrt{\frac{Var(\hat{h}_i)}{n}} = \sqrt{\frac{Var(h(x_i))+Var(h(y_i))+2Cov(h(x_i), h(y_i))}{4n}}$

若 $h(x_i)$ 和 $h(y_i)$ 是 negative 的. eg. h是 x y $h(x)\uparrow$ $h(y)\downarrow$

$SE_{new} \le \sqrt{\frac{2Var(h(x))}{4n}} = \sqrt{\frac{Var(h(x))}{2n}} \le SE_{old}$. $\quad Var(h(x_i)) = Var(h(x))$

若 positive x y $\underline{h(x)\uparrow \ h(y)\uparrow}$

$Cov(h(x_i), h(y_i)) \le \sqrt{Var(h(x_i))}\sqrt{Var(h(y_i))} = Var(X)$

$\therefore SE_{new} \le \sqrt{\frac{1}{4n}\left(Var(X)+Var(X)+2Var(X)\right)}$

$= \sqrt{\frac{1}{n}Var(X)} = SE_{old}$.

Finally, $SE_{new} \le SE_{old}$.

- Resampling and bootstrap
    - 为什么会有这个方法呢?
        - Scenarios 1：估计MSE，如果$\hat{\theta}_n$的表达式特别复杂很难算，我们没办法计算Var

$$MSE(\hat{\theta}_n) = Var(\hat{\theta}_n) + \underbrace{(\mathbb{E}[\hat{\theta}_n] - \theta)}_{Bia_\theta(\hat{\theta}_n)}{}^2.$$

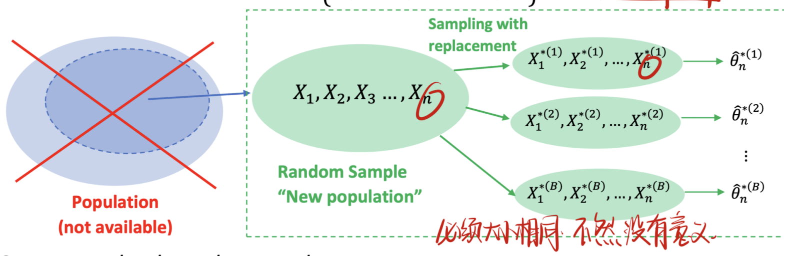
- Scenarios 2: 计算$\bar{X}$的置信区间，我们之前怎么算呢，$\bar{X}$服从于$N(\mu, Var(X)/n)$，但我们只是不知道Var(X),我们用样本方差代替Var(X)，从而$\frac{\bar{X}-\mu}{\sqrt{s^2/n}} = \frac{\frac{X-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{s^2/\sigma^2}} = \frac{Z}{\sqrt{卡方/(n-1)}}$服从于t(n-1)分布，但注意！！！！服从卡方分布是建立在X是正态分布的前提下，这样$\frac{s^2(n-1)}{\sigma^2}$才服从卡方分布，否则这个公式是不成立的，那怎么办呢？足够大的样本
- bootstrap
  - Procedure

    Procedure of bootstrapping

    Step 1: Given the sample data $\{X_1, X_2, \ldots, X_n\}$, a random sample of size n is randomly drawn *with replacement* from the sample.

    Step 2: We let $\{X_1^*, X_2^*, \ldots, X_n^*\}$ be the bootstrap sample obtained, we estimate $\hat{\theta}_n^{*(i)} = \hat{\theta}_n(X_1^*, \ldots, X_n^*)$, which is the bootstrap sample data of $\hat{\theta}_n$.

    Step 3: Repeated the procedure for sufficiently many times (say B times) and obtain a set of bootstrap sample $\left\{\hat{\theta}_n^{*(1)}, \hat{\theta}_n^{*(2)}, \ldots, \hat{\theta}_n^{*(B)}\right\}$. $\boxed{\text{resample}}$

    必须大小相同 不然没有意义.
  - Remark:
    - 一定要放回，不然每个X_i^*就不是iid了
    - 每次抽的样本大小一定要和原来的样本大小一样：我们用新的样本来估计原始样本，参数的大小和样本大小是相关的，所以为了估计n个原始数据的参数，我们抽取的样本也需要和原始样本的大小相同
      - 其实相当于把原来的sample data当作ture data 看待，然后从中抽取很多样本，每个样本获得一个θ，再将每个参数求平均，就是我们最后获得的参数；同样，如果我们想获得置信区间，只需要对这么多θ求分位数就可以了
      - Remark:
        - 1.error：
          - 1. simulation error
          - 2. sample distribution approximation error我们的参数估计的准确程度和我们的原始数据是相关的，如果给我们的原始数据对真实分布的估计比较准确，那么我们估计参数的准确率也就相对较高 $\rightarrow n$(每次样本的大小)越大，B(多少次重新抽样)越大

- parameter bootstrap
  - 建立在我们已知真实分布，但是真实分布存在一些参数，如何估计这些参数
  - MSE maximum， 如果我们的样本大小非常小，bootstrap非常有用来达到估计这些参数的目的
  - 但是现实生活中很难有真实分布已知的情况