Probability Inequality

Zha Mengyue & Huang Chutian

The Hong Kong University of Science and Technology

{*mzha, chuangat*}*@ust.hk*

February 24, 2022

Why we need them?

In machine learning, we want to know the

$$\mathbb{P}\{\text{the training error is close to the testing error}\}$$

We introduce a family of r.v's $\{X_i\}_{i=1}^n$ to describe the training error on n samples.

1. $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$: training error
2. $\mu = \mathbb{E}X_i$: the testing error

Bound the Deviation

Let $X_1, \ldots, X_n$ be n i.i.d r.v's, with expectation

$$\mu = \mathbb{E}X_i.$$

Let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Given $\epsilon > 0$, we are interested in estimating the following tail probabilities:

$$\Pr\left(\bar{X}_n \geq \mu + \epsilon\right)$$
$$\Pr\left(\bar{X}_n \leq \mu - \epsilon\right)$$

Now we would like to bound the probability of the event that empirical mean deviates significantly from the expectation.

Example 1: Standard Gaussian r.v's

Let $X \sim N(0, 1)$, with density function

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

For $\epsilon > 0$, we want to estimate the tail probability $\Pr(X \geq \epsilon)$ as follows.

$$\begin{aligned}
\Pr(X \geq \epsilon) &= \int_{\epsilon}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\
&= \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x+\epsilon)^2/2} dx \leq \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\left(x^2+\epsilon^2\right)/2} dx \\
&= 0.5 e^{-\epsilon^2/2}
\end{aligned}$$

More General Formula?

Theorem

Given any non-negative function $g(x) \geq 0$, we have for set S :

$$\Pr\left(\bar{X}_n \in S\right) \leq \frac{\mathbb{E}g\left(\bar{X}_n\right)}{\inf_{x \in S} g(x)}$$

Since $g(x)$ is non-negative, we have

$$\mathbb{E}g\left(\bar{X}_n\right) \geq \mathbb{E}_{\bar{X}_n \in S} g\left(\bar{X}_n\right) = \mathbb{E}[\mathbb{I}\{\bar{X}_n \in S\} g\left(\bar{X}_n\right)]$$

$$\geq g_S \mathbb{E}\mathbb{I}\{\bar{X}_n \in S\} = g_S \Pr\left(\bar{X}_n \in S\right),$$

where $g_s = \inf_{x \in S} g(x)$. This leads to the desired bound.

Example 1: $g(z) = (z - \mu)^2$

In particular, we consider the choice of $g(z) = (z - \mu)^2$ for Markov inequality.

$$\mathbb{E}g\left(\bar{X}_n\right) = \mathbb{E}\left(\bar{X}_n - \mu\right)^2 = \frac{1}{n}\operatorname{Var}\left(X_1\right)$$

Since $S = \left\{\left|\bar{X}_n - \mu\right| \geq \epsilon\right\}$, we have

$$\Pr\left(\left|\bar{X}_n - \mu\right| \geq \epsilon\right) \leq \frac{\operatorname{Var}\left(X_1\right)}{n\epsilon^2}$$

This is denoted as Chebyshev inequality.

Example 2: $g(z) = e^{\lambda n z}$

We choose $g(z) = e^{\lambda n z}$ with some tuning parameter $\lambda > 0$.

$$\begin{aligned}
\Pr\left(\bar{X}_n \geq \mu + \epsilon\right) &\leq \frac{\mathbb{E} g\left(\bar{X}_n\right)}{\inf_{x \in S} g(x)} = \frac{\mathbb{E} e^{\lambda n \bar{X}_n}}{e^{\lambda n(\mu+\epsilon)}} = \frac{\mathbb{E} e^{\lambda \sum_{i=1}^n X_i}}{e^{\lambda n(\mu+\epsilon)}} \\
&= \frac{\mathbb{E} \prod_{i=1}^n e^{\lambda X_i}}{e^{\lambda n(\mu+\epsilon)}} = e^{-\lambda n(\mu+\epsilon)} \left[\mathbb{E} e^{\lambda X_1}\right]^n
\end{aligned}$$

Therefore by taking logarithm, we obtain

$$\ln \Pr\left(\bar{X}_n \geq \mu + \epsilon\right) \leq n\left[-\lambda(\mu+\epsilon) + \ln \mathbb{E} e^{\lambda X_1}\right]$$

Example 2: $g(z) = e^{\lambda n z}$

Taking infimum over $\lambda > 0$ on the RHS, we have

$$n^{-1} \ln \Pr\left(\bar{X}_n \geq \mu + \epsilon\right) \leq \inf_{\lambda > 0} \left[-\lambda\epsilon + \ln \mathbb{E} e^{\lambda(X_1 - \mu)}\right]$$

Similarly,

$$n^{-1} \ln \Pr\left(\bar{X}_n \leq \mu - \epsilon\right) \leq \inf_{\lambda < 0} \left[\lambda\epsilon + \ln \mathbb{E} e^{\lambda(X_1 - \mu)}\right]$$

To simplify the formula, we denote the rate function $I_X(z)$ as

$$-I_X(z) = \inf_{\lambda > 0} \left[-\lambda z + \ln \mathbb{E}_X e^{\lambda X}\right]$$

Exponential Tail Inequality

Theorem

$$\frac{1}{n}\ln\Pr\left(\bar{X}_n \geq \mu+\epsilon\right) \leq -I_{X_1}(\mu+\epsilon) = \inf_{\lambda>0}\left[-\lambda\epsilon + \ln\mathbb{E}e^{\lambda(X_1-\mu)}\right]$$

$$\frac{1}{n}\ln\Pr\left(\bar{X}_n \leq \mu-\epsilon\right) \leq -I_{X_1}(\mu-\epsilon) = \inf_{\lambda<0}\left[\lambda\epsilon + \ln\mathbb{E}e^{\lambda(X_1-\mu)}\right]$$

We define the logarithmic moment generating function of the random variable X,

$$\Gamma_X(\lambda) = \ln\mathbb{E}e^{\lambda X}.$$

Then,

$$-I_X(z) = \inf_{\lambda>0}\left[-\lambda z + \Gamma_X(\lambda)\right]$$

Later, we bound the rate function $-I_X(z)$ by different techniques.

Example: Gaussian r.v's

Assume that $X_i \sim N\left(\mu, \sigma^2\right)$, then the exponential moment is

$$\begin{aligned}\mathbb{E}e^{\lambda(X_1-\mu)} &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\lambda x} e^{-x^2/2\sigma^2} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\lambda^2\sigma^2/2} e^{-(x/\sigma-\lambda\sigma)^2/2} dx/\sigma = e^{\lambda^2\sigma^2/2}\end{aligned}$$

Therefore,

$$-I_{X_1}(\mu+\epsilon) = \inf_{\lambda>0}\left[-\lambda\epsilon + \ln \mathbb{E}e^{\lambda(X_1-\mu)}\right] = \inf_{\lambda>0}\left[-\lambda\epsilon + \frac{\lambda^2\sigma^2}{2}\right] = -\frac{\epsilon^2}{2\sigma^2},$$

where the optimal λ is achieved at $\lambda = \epsilon/\sigma^2$.

Bound $-I_X(z)$ for different X

Now we discuss the $-I_X(z)$ for X with different conditions

1. sub-gaussian
2. bounded (Hoeffding Inequality)
3. upper bounded + small variance (Bernnett Inequality)
4. moment conditions (Berstein Inequality)

Sub-Gaussian r.v's

Definition

A sub-Gaussian random variable X has quadratic logarithmic moment generating function:

$$\ln \mathbb{E} e^{\lambda X} \leq \lambda\mu + \frac{\lambda^2}{2} b.$$

Then for any $z > \mu$:

$$-I_X(z) = \inf_{\lambda > 0} \left(-\lambda z + \lambda\mu + \frac{\lambda^2}{2} b \right).$$

taking derivative w.r.t λ, the optimal $\lambda_* = (\mu - z)/b$. Thus,

$$I_X(z) = \frac{(z-\mu)^2}{2b}.$$

Bound the $\Gamma_X(\lambda)$

Lemma

Consider r.v. X with $\mathbb{E}[X] = \mu$. If $\exists \alpha > 0, \beta \geq 0$ s.t. for $\lambda \in \left[0, \beta^{-1}\right)$,

$$\Gamma_X(\lambda) \leq \lambda\mu + \frac{\alpha\lambda^2}{2(1-\beta\lambda)},$$

then for $\epsilon > 0$:

$$-I_X(\mu+\epsilon) \leq -\frac{\epsilon^2}{2(\alpha+\beta\epsilon)}.$$

Theorem

If X has a $\Gamma_X(\lambda)$ satisfies the lemma above for $\lambda > 0$, then all $\epsilon > 0$:

$$\Pr\left(\bar{X}_n \geq \mu+\epsilon\right) \leq \exp\left[\frac{-n\epsilon^2}{2(\alpha+\beta\epsilon)}\right]$$

Tail Inequality for Sub-Gaussians

Our condition in the lemma is

$$\Gamma_X(\lambda) \leq \lambda\mu + \frac{\alpha\lambda^2}{2(1-\beta\lambda)}$$

Let $\alpha = b$ and $\beta = 0$, We have

$$\Gamma_X(\lambda) \leq \lambda\mu + \frac{\lambda^2}{2}b$$

We therefore reach the the inequality for sub-gaussians by $\alpha = b$ and $\beta = 0$.

Theorem

If X is sub-Gaussian, then for all $\epsilon > 0$:

$$\Pr\left(\bar{X}_n \geq \mu + \epsilon\right) \leq e^{-n\epsilon^2/2b}$$
$$\Pr\left(\bar{X}_n \leq \mu - \epsilon\right) \leq e^{-n\epsilon^2/2b}$$

Bounded r.v. (Hoeffding's Inequality/ Chernoff Bound)

Hoeffding's inequality is an exponential tail inequality for **bounded** random variables. It is also referred to as Chernoff bound.

Lemma

Consider a random variable $X \in [0,1]$ and $\mathbb{E}X = \mu$. We have the following inequality

$$\mathbb{E}e^{\lambda X} \leq (1-\mu)e^0 + \mu e^\lambda$$

Moreover,

$$\ln\left[(1-\mu)e^0 + \mu e^\lambda\right] \leq \lambda\mu + \lambda^2/8$$

Proof

Proof.

Let $h_L(\lambda) = Ee^{\lambda X}$ and $h_R(\lambda) = (1-\mu)e^0 + \mu e^\lambda$. We have $h_L(0) = h_R(0)$. Moreover, $h'_L(\lambda) = \mathbf{EX}e^{\lambda X} \leq \mathbf{EX}e^{\lambda * 1} = \mu e^\lambda = h'_R(\lambda)$. This proves the first inequality. For the second inequality, let

$$h(\lambda) = \ln\left[(1-\mu)e^0 + \mu e^\lambda\right]$$

then

$$h'(\lambda) = \frac{\mu e^\lambda}{(1-\mu)e^0 + \mu e^\lambda}$$

and

$$h''(\lambda) = \frac{\mu e^\lambda}{(1-\mu)e^0 + \mu e^\lambda} - \frac{\left(\mu e^\lambda\right)^2}{[(1-\mu)e^0 + \mu e^\lambda]^2} = \left|h'(\lambda)\right|\left(1 - \left|h'(\lambda)\right|\right) \leq 1/4$$

By Taylor expansion, we obtain $h(\lambda) \leq h(0) + \lambda h'(0) + \lambda^2/8$. □

Additive Forms of Chernoff Bounds

In this way, we have the bound of $\Gamma_X(\lambda)$ for a bounded random variable:

$$\Gamma_X(\lambda) \leq \lambda\mu + \lambda^2/8,$$

which corresponds to the case of sub-Gaussian r. v. by taking $b = 1/4$.

Theorem (Additive Forms of Chernoff Bounds)

Assume that $X_1 \in [0, 1]$. Then for all $\epsilon > 0$:

$$\Pr\left(\bar{X}_n \geq \mu + \epsilon\right) \leq e^{-2n\epsilon^2}$$
$$\Pr\left(\bar{X}_n \leq \mu - \epsilon\right) \leq e^{-2n\epsilon^2}$$

Bound the rate function by KL Divergence

In some applications, one often needs to employ a more refined form of Chernoff bound. We consider bounding the rate function by KL divergence. Consider the case $z = \mu + \epsilon$. We have

$$-I_{X_1}(z) \leq \inf_{\lambda>0} \left[-\lambda z + \ln \left((1-\mu)e^0 + \mu e^\lambda \right) \right].$$

Assume that the optimal value of λ on the right hand side optimum is achieved at λ_*. By setting the derivative to zero, we obtain the expression:

$$z = \frac{\mu e^{\lambda_*}}{(1-\mu)e^0 + \mu e^{\lambda_*}} \implies e^{\lambda_*} = \frac{z(1-\mu)}{\mu(1-z)}$$

Substituting λ_* to the RHS, we have $-I_{X_1}(z) \leq -\mathrm{KL}(z\|\mu)$.
The case of $z = \mu - \epsilon$ is similar.

Tail Inequality for bounded r. v. with KL Divergence

Theorem

Assume that $X_1 \in [0,1]$. Then for all $\epsilon > 0$, we have

$$\Pr\left(\bar{X}_n \geq \mu + \epsilon\right) \leq e^{-n\mathrm{KL}(\mu+\epsilon\|\mu)}$$
$$\Pr\left(\bar{X}_n \leq \mu - \epsilon\right) \leq e^{-n\mathrm{KL}(\mu-\epsilon\|\mu)}$$

where $\mathrm{KL}(z\|\mu)$ is the Kullback-Leibler divergence (KL divergence) defined as

$$\mathrm{KL}(z\|\mu) = z \ln \frac{z}{\mu} + (1-z) \ln \frac{1-z}{1-\mu}.$$

Multiplicative Form of Chernoff Bounds

Theorem (Multiplicative Form of Chernoff Bounds)

Assume that $X_1 \in [0,1]$. Then for all $\epsilon > 0$:

$$\Pr\left(\bar{X}_n \geq (1+\epsilon)\mu\right) \leq \exp\left[\frac{-n\mu\epsilon^2}{2+\epsilon}\right]$$

$$\Pr\left(\bar{X}_n \leq (1-\epsilon)\mu\right) \leq \exp\left[\frac{-n\mu\epsilon^2}{2}\right]$$

Moreover, for $t > 0$, we have:

$$\Pr\left(\bar{X}_n \geq \mu + \sqrt{\frac{2\mu t}{n}} + \frac{t}{3n}\right) \leq e^{-t}$$

Proof: $\mathrm{KL}(z\|\mu) \geq (z-\mu)^2 / \max(2\mu, \mu+z)$, $\Lambda_{X_1}(\lambda) \leq \ln\left[(1-\mu)e^0 + \mu e^\lambda\right] \leq \mu\left(e^\lambda - 1\right) = \mu(\lambda + \sum_{k\geq 2}\frac{\lambda^k}{k!}) \leq \mu\lambda + \frac{\mu\lambda^2}{2(1-\lambda/3)}$.

Bennett's Inequality

In Bennett's inequality, we assume that the random variable is **upper bounded**, and has a **small variance**.

Lemma

If $X - \mathbb{E}X \leq b$, *then* $\forall \lambda \geq 0$ *:*

$$\ln \mathbb{E}e^{\lambda X} \leq \lambda \mathbb{E}X + \lambda^2 \phi(\lambda b) \mathbb{V}(X)$$

where $\phi(z) = \left(e^z - z - 1\right)/z^2$.

Bennett's Inequality

Proof: Let $X' = X - \mathbb{E}X$. We have

$$\begin{aligned}
\ln \mathbb{E}e^{\lambda X} &= \lambda \mathbb{E}X + \ln Ee^{\lambda X'} \\
&\leq \lambda \mathbb{E}X + \mathbb{E}e^{\lambda X'} - 1 \\
&= \lambda \mathbb{E}X + \lambda^2 \mathbb{E} \frac{e^{\lambda X'} - \lambda X' - 1}{(\lambda X')^2} \left(X'\right)^2 \\
&\leq \lambda \mathbb{E}X + \lambda^2 \mathbb{E}\phi(\lambda b) \left(X'\right)^2 .
\end{aligned}$$

$$-I_{X_1}(\mu + \epsilon) \leq \inf_{\lambda > 0} \left[-\lambda \epsilon + b^{-2} \left(e^{\lambda b} - \lambda b - 1 \right) \mathbb{V}(X_1) \right],$$

Set the derivative of the objective function (RHS without inf) w.r.t. λ to zero, and obtain the optimal λ satifies:

$$-\epsilon + b^{-1} \left(e^{\lambda b} - 1 \right) \mathbb{V}(X_1) = 0$$

Bennett's Inequality

This gives $\lambda = b^{-1} \ln(1 + \epsilon b/\mathbb{V}(X_1))$. Plugging it in RHS,

$$-I_{X_1}(\mu + \epsilon) \leq -\frac{\mathbb{V}(X_1)}{b^2} \psi \left(\frac{\epsilon b}{\mathbb{V}(X_1)} \right).$$

Given $\lambda \in (0, 3/b)$, by Taylor expansion of the exponential function,

$$\begin{aligned}
\Lambda_{X_1}(\lambda) &\leq \mu\lambda + b^{-2} \left[e^{\lambda b} - \lambda b - 1 \right] \mathbb{V}(X_1) \\
&\leq \mu\lambda + \frac{\mathbb{V}(X_1)\lambda^2}{2} \sum_{m=0}^{\infty} (\lambda b/3)^m \\
&= \mu\lambda + \frac{\mathbb{V}(X_1)\lambda^2}{2(1 - \lambda b/3)}.
\end{aligned}$$

Bennett's Inequality

Theorem (Bennett's Inequality)

If $X_1 \le \mu + b$, *for some* $b > 0$. *Let* $\psi(z) = (1+z)\ln(1+z) - z$, *then* $\forall \epsilon > 0$ *:*

$$\Pr\left[\bar{X}_n \ge \mu + \epsilon\right] \le \exp\left[\frac{-n\mathbb{V}(X)}{b^2}\psi\left(\frac{\epsilon b}{\mathbb{V}(X)}\right)\right]$$

$$\Pr\left[\bar{X}_n \ge \mu + \epsilon\right] \le \exp\left[\frac{-n\epsilon^2}{2\mathbb{V}\left(X_1\right) + 2\epsilon b/3}\right]$$

Moreover, for $t > 0$ *:*

$$\Pr\left[\bar{X}_n \ge \mu + \sqrt{\frac{2\mathbb{V}\left(X_1\right)t}{n}} + \frac{bt}{3n}\right] \le e^{-t}$$

In Bernstein's inequality, we obtain results similar to the Bennett's inequality, but using a **moment condition** instead of the boundedness condition.

Lemma

If X satisfies the following moment condition with $b, V > 0$ for integers $m \geq 2$:

$$\mathbb{E}[X - c]^m \leq m!(b/3)^{m-2} V/2,$$

where c is arbitrary. Then when $\lambda \in (0, 3/b)$:

$$\ln \mathbb{E}e^{\lambda X} \leq \lambda \mathbb{E}X + \frac{\lambda^2 V}{2(1 - \lambda b/3)}$$

Proof:

$$\ln \mathbb{E}e^{\lambda X} \leq \lambda c + \mathbb{E}e^{\lambda(X-c)} - 1 \leq \lambda \mathbb{E}X + 0.5V\lambda^2 \sum_{m=2} (b/3)^{m-2}\lambda^{m-2}$$

$$= \lambda \mathbb{E}X + 0.5\lambda^2 V(1 - \lambda b/3)^{-1}.$$

Bernstein's Inequality

Theorem (Bernstein's Inequality)

Assume that X_1 satisfies the moment condition in the above lemma. Then for all $\epsilon > 0$:

$$\Pr\left[\bar{X}_n \geq \mu + \epsilon\right] \leq \exp\left[\frac{-n\epsilon^2}{2V + 2\epsilon b/3}\right]$$

and for all $t > 0$:

$$\Pr\left[\bar{X}_n \geq \mu + \sqrt{\frac{2Vt}{n}} + \frac{bt}{3n}\right] \leq e^{-t}.$$

Example

If the random variable X is bounded with $|X - \mu| \leq b$, then the moment condition lemma holds with $c = \mu$ and $V = \mathbb{V}(X)$.

Tail Inequality for χ^2 Distribution

Definition (χ_n)

Let $X_i \sim N(0, 1)$ be IID normal random variables ($i = 1, \ldots, n$), then the random variable

$$Z = \sum_{i=1}^{n} X_i^2$$

is distributed according to the chi-square distribution with n degrees of freedom, denoted by χ_n. This random variable plays an important role in the analysis of least squares regression with Gaussian noise.

Tail Inequality for χ^2 Distribution

Theorem

Let $Z \sim \chi_n$, then for all $t > 0$:

$$\Pr[Z \geq n + 2\sqrt{nt} + 2t] \leq e^{-t}$$

and

$$\Pr[Z \leq n - 2\sqrt{nt}] \leq e^{-t}$$

Proof.

The logarithmic moment generating function of X_1^2 for $\lambda \in [0, 0.5)$ is

$$\Lambda_{X_1^2}(\lambda) = \ln \mathbb{E} e^{\lambda X_1^2} = -0.5 \ln(1 - 2\lambda) = 0.5 \sum_{k=1}^{\infty} \frac{(2\lambda)^k}{k}$$

$$\leq \lambda + \lambda^2 \sum_{k \geq 0} (2\lambda)^k = \lambda + \frac{\lambda^2}{1 - 2\lambda}.$$

Proof. Cont. (Tail Inequality for χ^2 Distribution)

Proof. Cont.

The first equality can be obtained using Gaussian integration. The second equality used the Taylor expansion of the logarithm function. Note that $\mu = \mathbb{E}X_1^2 = 1$.

The first bound of the theorem follows from theorem in page 13 with $\alpha = 2$ and $\beta = 2$.

If $\lambda \leq 0$, then

$$\Lambda_{X_1^2}(\lambda) = \ln \mathbb{E}e^{\lambda X_1^2} = -0.5 \ln(1 - 2\lambda) \leq \lambda + \lambda^2.$$

The second inequality now follows from the sub-Gaussian tail inequality with $b = 2$. □

Non-identically Distributed Random Variables

If $X_1, \ldots, X_n$ are independent but not identically distributed random variables, then a tail inequality similar to that of i.i.d. r.v. holds. Let $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $\mu = \mathbb{E}\bar{X}_n$, then we have the following bound.

Theorem

We have

$$\Pr\left(\bar{X}_n \geq \mu + \epsilon\right) \leq \inf_{\lambda > 0}\left[-\lambda n(\mu+\epsilon) + \sum_{i=1}^{n} \ln \mathbb{E} e^{\lambda X_i}\right]$$

Corollary

If X_i are sub-Gaussians with $\ln \mathbb{E} e^{\lambda X_i} \leq \lambda \mathbb{E} X_i + 0.5\lambda^2 b_i$, then for all $\epsilon > 0$:

$$\Pr\left(\bar{X}_n \geq \mu + \epsilon\right) \leq \exp\left[-\frac{n^2\epsilon^2}{2\sum_{i=1}^n b_i}\right]$$

Non-identically Distributed Random Variables

Corollary

Let $\sigma_i = \{\pm 1\}$ be independent random Bernoulli variables (each takes value ± 1 with equal probability). Let a_i be fixed numbers ($i = 1, \ldots, n$). Then

$$\Pr\left(n^{-1}\sum_{i=1}^{n}\sigma_i a_i \geq \epsilon\right) \leq \exp\left[-\frac{n\epsilon^2}{2n^{-1}\sum_{i=1}^{n} a_i^2}\right].$$

Corollary

If $X_i - \mathbb{E}X_i \leq b$ for all i, then

$$\Pr\left(\bar{X}_n \geq \mu + \epsilon\right) \leq \exp\left[-\frac{n^2\epsilon^2}{2\sum_{i=1}^{n}\mathbb{V}\left(X_i\right) + 2nb\epsilon/3}\right]$$

A more often preferred form in the theoretical analysis of machine learning

In the sub-Gaussian tail bound, we have for any $t > 0$,

$$\Pr\left(\bar{X}_n \geq \mu + \sqrt{\frac{2bt}{n}}\right) \leq e^{-t}.$$

Consider $\delta \in (0,1)$ such that $\delta = \exp(-t)$, we have $t = \ln(1/\delta)$. Then the inequality can be rewrited as:
With probability at least $1-\delta$, we have

$$\bar{X}_n < \mu + \sqrt{\frac{2b\ln(1/\delta)}{n}}.$$

Similarly, the Hoffding's inequality (Chernoff bound), Bennett's inequality, Bernstein's inequality can also be rewrited in this way.

Conclusions

1. Exponential Tail Inequalites can be used to bound the difference of true mean and the observed empirical mean.
2. r.v. is sub-Gaussian: with deviation of order $O(1/\sqrt{n})$
3. r.v is bounded: a special case of sub-Gaussian: Chernoff bound
4. r.v. is upper bounded and has a small variance: Bennett's inequality
5. r.v. has a certain moment condition: Bernstein's inequality
6. χ^2 distribution, non-identically distributed distribution....

Thanks!