Log-likelihood values and Monte Carlo simulation - some fundamental results

Land, Ingmar; Hoeher, Peter; Sorger, Ulrich

Published in:
Ikke angivet

Publication date:
2000

Document Version
Publisher's PDF, also known as Version of record

Log-Likelihood Values and Monte Carlo Simulation – Some Fundamental Results

Peter Hoeher and Ingmar Land

Information and Coding Theory Lab
University of Kiel, Germany
{ph,il}@tf.uni-kiel.de

Ulrich Sorger

Institute for Communications Technology
Darmstadt University of Technology, Germany
uli@nesi.tu-darmstadt.de

Abstract: *The purpose of this paper is twofold: We derive some fundamental properties of log-likelihood ratio (LLR) values and propose two novel "soft-decision" Monte Carlo simulation techniques based on probabilities or LLR values. Specifically, we prove that the pdf of LLR values is exponential symmetric and demonstrate that soft-decision simulation outperforms conventional bit error rate simulations with respect to accuracy and/or simulation time.*

Keywords: Monte Carlo simulation, log-likelihood values, a posteriori probability decoding.

1. INTRODUCTION

In the context of iterative decoding, soft-input soft-output decoders play an important role. A suitable component decoder is the a posteriori probability (APP) decoder, which is optimal in terms of minimizing the bit error rate in the presence of a single code, and which provides close to optimal results in conjunction with iterative decoding in the presence of concatenated codes [1], [2]. The APP decoder may output probabilities, see for example [3], [1], [2], or LLR values, see for example [4], among other equivalent alternatives. The use of LLR values offers practical advantages, such as numerical stability, but also provides theoretical insights ("decoding $\hat{=}$ LLR amplification" [5]). The aim of this paper is to explore further, previously unpublished, fundamental properties of LLR values and to use them in the Monte Carlo simulation of rare error events.

We assume binary transmission and the existence of an APP decoder (which implies known channel statistics). The main results are as follows:

(1) The average bit error rate (BER) can be written as $P_b = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} P_{b,k}$, where $P_{b,k}$ is a single estimate of the average BER and K the number of estimates, i.e. the number of transmitted info bits. In classical Monte Carlo BER simulation (see 2.1., Method 1), $P_{b,k}$ takes on the values 1 for "error" and 0 for "no error", whereas in soft-decision simulation (see 2.1., Method 2), $P_{b,k} = \frac{1}{1+\exp\{|L_k|\}}$ is used. As opposed to the classical simulation where errors are counted (i.e., only the signs are evaluated), this formula suggests to use the APPs or the absolute values of the LLR values, respectively.

(2) The pdf of LLR values $L_k \in \mathbb{R}$ is exponential-symmetric: $p(L_k) = \exp\{L_k\} \cdot p(-L_k)$. This results holds for any APP decoder and channel. The exponential-symmetric property is not only interesting from a theoretical point of view but also from a practical, since it can be exploited for further improvement of BER simulation (see 2.1., Method 3).

2. BIT ERROR PROBABILITY AND LOG-LIKELIHOOD RATIOS

For the simulation of the BER, we use the setup depicted in Figure 1. The info bits $U \in \{+1, -1\}$ are channel encoded (ENC) and transmitted over a discrete memoryless channel (DMC). The received values $Y \in \mathbb{R}$ are fed into a channel decoder (DEC) which computes the a posteriori LLRs $L \in \mathbb{R}$ of the info bits as

$$\begin{aligned} L \hat{=} L(U|y) &= \log \frac{P_U(+1|y)}{P_U(-1|y)} = \log \frac{P_U(+1|y)}{1 - P_U(+1|y)} \\ &= \underbrace{\log \frac{P_U(+1)}{P_U(-1)}}_{\text{a priori information}} + \underbrace{\log \frac{p_Y(y|+1)}{p_Y(y|-1)}}_{\text{channel information}} \quad . \end{aligned} \tag{1}$$

For an efficient implementation, the so-called LogAPP algorithm (symbol-by-symbol LogMAP algorithm [4], forward-backward algorithm, Log-BCJR algorithm) can be applied. Finally, the sent info bits are estimated according to the sign of their respective LLR.

The average bit error rate P_b of the transmission system can be estimated in (at least) three different ways.

2.1. BER Simulation Methods

2.1.1. Method 1

Compare the sign of the info bit U and the sign of its LLR L at the decoder output. If they agree, set the bit error indicator E to 0, otherwise set it to 1:

$$E = \begin{cases} 0 & \text{if} \quad \operatorname{sgn}(U) = \operatorname{sgn}(L), \\ 1 & \text{else.} \end{cases}$$

The BER is the expected value[1] of the random variable E:

$$P_b = \mathop{\mathrm{E}}_{e}[E] = \sum_{e} P_E(e)\, e \quad , \tag{2}$$

[1]The expected value is denoted by E [.] throughout this paper.

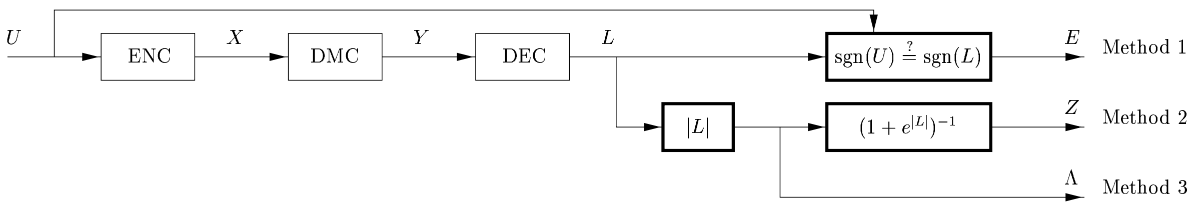

Figure 1: Simulation setup.

with $e \in \{0,1\}$ denoting realizations of E. When Method 1 is applied to a sample of K transmitted bits, an estimate of the BER is $\hat{P}_b = \frac{1}{K}\sum_{k=1}^{K} E_k$.

2.1.2. Method 2

Take the absolute value $\Lambda = |L|$ of the LLR L and compute Z as

$$Z = \frac{1}{1+e^{|L|}} = \frac{1}{1+e^{\Lambda}} \quad .$$

This is the probability that the hard decision of the info bit is wrong, i.e. that the signs of U and L differ. This becomes obvious, when (1) is expressed as

$$\begin{aligned} P_U(+1|y) &= \frac{1}{1+e^{-L}} = \frac{1}{1+e^{(-1)\Lambda}} \quad , \\ P_U(-1|y) &= \frac{1}{1+e^{+L}} = \frac{1}{1+e^{(+1)\Lambda}} \quad . \end{aligned}$$

Due to this property, Z can be regarded as a "soft" bit error indicator. Like in the previous method, the BER is the expected value of Z:

$$P_b = \mathop{\mathrm{E}}_z[Z] = \int_z p_Z(z)\; z\; dz \quad , \tag{3}$$

with $z \geq 0$ denoting realizations of the random variable Z. When Method 2 is applied to a sample of K transmitted bits, an estimate of the BER is $\hat{P}_b = \frac{1}{K}\sum_{k=1}^{K} Z_k$.

2.1.3. Method 3

Take the absolute value $\Lambda = |L|$ of the LLR L. Then, the BER is the estimated value of a function of Λ, namely

$$P_b = \mathop{\mathrm{E}}_\lambda\left[\frac{1}{1+e^{\lambda}}\right] = \int_\lambda p_\Lambda(\lambda)\;\frac{1}{1+e^{\lambda}}\; d\lambda \quad , \tag{4}$$

with $\lambda \geq 0$ denoting realizations of the random variable Λ. This method efficiently utilizes the exponential-symmetric property of the LLR (see 3.3.). When Method 3 is applied to a sample of K transmitted bits, an estimate $\hat{p}_\Lambda(\lambda)$ of $p_\Lambda(\lambda)$ must be computed (e.g. by means of a histogram), before the integral in (4) can be solved numerically to obtain an estimate $\hat{P}_b$ of the BER. This corresponds to block processing.

Comparison: For each of these methods, the task of estimating the BER is equivalently to the task of estimating the probability distribution of the respective random variable. Since $p_\Lambda(\lambda)$ typically is smoother than $P_E(e)$ and $p_Z(z)$, this might be more efficient for Λ (Method 3) than for E (Method 1) and Z (Method 2).

Whereas in classical simulation (Method 1) the sent info bits have to be known, this additional information is *not* necessary in "soft-decision" simulation (Method 2 and Method 3). Since the latter two methods do not need the information "phase" (the sign), they could be denoted as "incoherent simulation". Correspondingly, the first method could be denoted as "coherent simulation". Note the practical advantage of "incoherent simulation".

2.2. Analysis and Evaluation of the BER Simulation Methods

When Method 1 or Method 2 is applied, the estimated BER is the mean of K random variables. Therefore, the variance of $\hat{P}_b$ is given by σ_E^2/K or σ_Z^2/K, respectively. Since the variances are appropriate means to compare the quality of these two methods, they are discussed in the following.

2.2.1. Variance of Method 1

The probability distribution of the "hard" bit error indicator E is given by

$$P_E(e) = \begin{cases} 1-P_b & \forall\; e=0 \\ P_b & \forall\; e=1\;. \end{cases}$$

Thus, the variance of this binary random variable computes as

$$\sigma_E^2 = \mathop{\mathrm{E}}_e[E^2] - (\mathop{\mathrm{E}}_e[E])^2 = P_b\cdot(1-P_b) \quad . \tag{5}$$

2.2.2. Variance of Method 2

The computation of σ_Z^2 is a bit more complicated. Firstly, the pdf of Z is expressed by means of the pdf of Λ. Since $Z = \left(1+e^{\Lambda}\right)^{-1}$, or equivalently $\Lambda = \ln\frac{1-Z}{Z}$, the equation

$$p_Z(z) = \frac{1}{z(1-z)}\cdot p_\Lambda\Big(\ln\frac{1-z}{z}\Big)$$

can be derived by a simple variable substitution argument. Taking (11) into consideration, the pdf of Z can be formulated by means of the conditioned pdf of L:

$$p_Z(z) \;=\; \frac{1}{z(1-z)^2} \cdot p_L\Big(\ln \frac{1-z}{z} \;\Big|\; +1\Big) \;\; .$$

Given this result, the variance of this positive, real-valued random variable computes as

$$\sigma_Z^2 = \mathop{\mathrm{E}}_z[Z^2] - (\mathop{\mathrm{E}}_z[Z])^2 \;\; . \qquad (6)$$

2.3. Applications

Uncoded system: For uncoded transmission with BPSK over an AWGN channel with noise variance σ_n^2, $P_b = \mathrm{Q}\,(1/\sigma_n)$ and the variance of E results in $\sigma_E^2 = \mathrm{Q}\,(1/\sigma_n) \cdot [1 - \mathrm{Q}\,(1/\sigma_n)]$ [2]. The variance of Z can not be reasonably simplified; but since $p_Z(z)$ is given analytically, σ_Z^2 can evaluated numerically. In Figure 2, the standard deviations of E and Z are plotted versus the channel SNR. The curves show that the standard deviations differ by about a factor of 2. Since the standard deviation of $\hat{P}_b$ decreases with $\sqrt{K}$, an estimation with Method 1 needs four times more samples K than an estimation with Method 2 to guarantee the same accuracy, i.e. the same standard deviation.

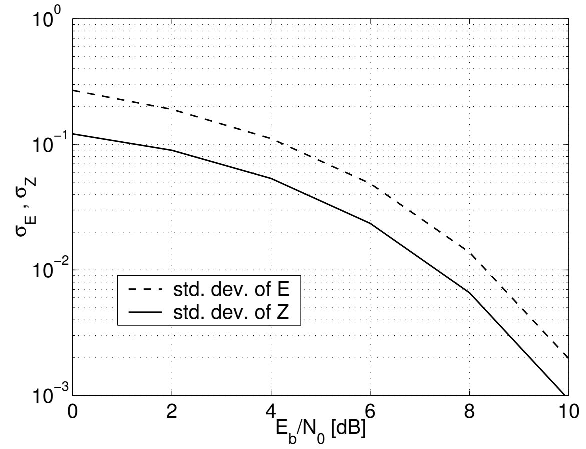

Figure 2: Standard deviations of E and Z for uncoded transmission over an AWGN channel.

Coded system: The superiority of Method 2 is demonstrated in Fig. 3 for transmission on the AWGN channel using a memory 3 convolutional code and APP decoding. The desired BER is $P_b = 10^{-4}$. The plot illustrates the standard deviation for both Method 1 (classical simulation) and Method 2 (the soft-decision simulation). Furthermore, the average BER for two specific realizations is shown. Again, the standard deviations differ by a factor of about 2.

[2] $\mathrm{Q}\,(x) \triangleq 1/\sqrt{2\pi} \int_x^\infty \exp(-t^2/2)dt$

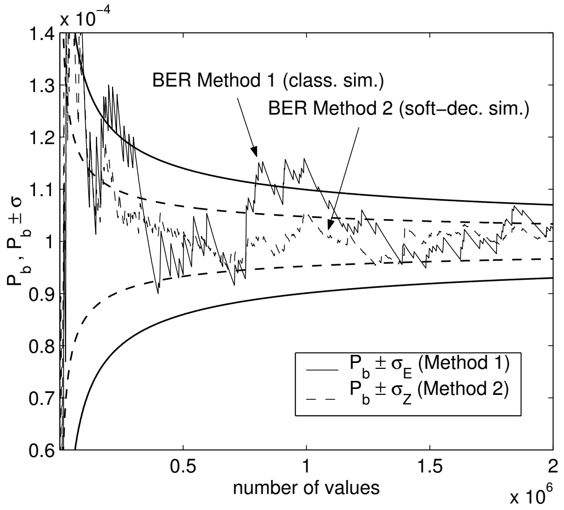

Figure 3: Comparison of different BER simulation techniques for convolutionally coded binary signaling on the AWGN channel.

3. PROPERTIES OF LOG-LIKELIHOOD RATIOS

3.1. Exponential Symmetry

The conditioned LLR of the binary random variable U is defined in (1). When the LLR is regarded as a random variable L with realization l and when $L(U) = 0$ is assumed, i.e., when the input U is uniformly distributed[3], then

$$p_Y(y|+1) \;=\; e^l p_Y(y|-1) \qquad (7)$$

follows from the definition of the LLR. For a transmission system comprising a symmetric channel and a linear channel code,

$$\boxed{\begin{aligned} p_L(l|+1) &= p_L(-l|-1) \\ p_L(l|-1) &= p_L(-l|+1) \;\; . \end{aligned}} \qquad (8)$$

Combining (7) and (8), a special property of the conditioned pdf of the LLR L becomes obvious:

$$\begin{aligned} p_L(l|+1) &= \int_{y:l} p_Y(y|+1)\, dy \\ &\overset{(7)}{=} \int_{y:l} e^l p_Y(y|-1)\, dy \\ &= e^l p_L(l|-1) \;\; , \end{aligned} \qquad (9)$$

where the integration is to be taken over all y that lead to the LLR l according to (1). This *exponential-symmetric property* is summarized in the following equation chain:

$$\boxed{\begin{aligned} p_L(l|+1) &\overset{(9)}{=} e^l \cdot p_L(l|-1) \\ &\overset{(8)}{=} e^l \cdot p_L(-l|+1) \;\; . \end{aligned}} \qquad (10)$$

[3] A generalization is possible.

Equations (8) and (10) describe two fundamental properties of the conditioned pdf of the LLR of a binary random variable that is transmitted over a linear, symmetric channel (coded or uncoded).

Applying this property, the relation between the conditioned pdf $p_L(l|+1)$ of the LLR L and the pdf $p_\Lambda(\lambda)$ of the LLR's absolute value Λ can be derived.

Firstly, the pdf $p_L(l)$ can be expressed by the conditioned pdf of L as:

$$\begin{aligned} p_L(l) &= P_U(+1)p_L(l|+1) + P_U(-1)p_L(l|-1) \\ &\overset{(10)}{=} P_U(+1)p_L(l|+1) + P_U(-1)e^{-l}p_L(l|+1) \\ &= \big[P_U(+1) + e^{-l}P_U(-1)\big]p_L(l|+1) \end{aligned}$$

and

$$\begin{aligned} p_L(-l) &= P_U(+1)p_L(-l|+1) + P_U(-1)p_L(-l|-1) \\ &\overset{(8),(10)}{=} P_U(+1)e^{-l}p_L(l|+1) + P_U(-1)p_L(l|+1) \\ &= \big[e^{-l}P_U(+1) + P_U(-1)\big]p_L(l|+1) \ . \end{aligned}$$

Then, the pdf $p_\Lambda(l)$ of the absolute value Λ can be computed as

$$\begin{aligned} p_\Lambda(l) &= p_L(l) + p_L(-l) \\ &= \underbrace{[P_U(+1) + P_U(-1)]}_{1}[1 + e^{-l}]p_L(l|+1) \ , \end{aligned}$$

which is valid for all $l \geq 0$. For $l < 0$, $p_\Lambda(l)$ is obviously equal to zero. Thus, the conditioned pdf $p_L(l|+1)$ of L and the pdf $p_\Lambda(l)$ of its absolute value Λ are linked by

$$\boxed{p_L(l|+1) = \frac{p_\Lambda(l)}{1+e^{-l}} \ \ \forall\, l \geq 0 \ .} \qquad (11)$$

3.2. Example for Exponential-Symmetric Distribution

As an example, let's look at the LLR of an uncoded BPSK transmission over an AWGN channel with the one-sided noise power density N_0, i.e., with noise variance $\sigma_n^2 = N_0/2$. For the fixed input $U = +u,\ u > 0$, the output Y is Gaussian distributed with mean value $\mu_Y = +u$ and variance $\sigma_Y^2 = \sigma_n^2 = N_0/2$, i.e.

$$p_Y(y|+1) = \frac{1}{\sqrt{2\pi}\,\sigma_Y}\ \exp\Big(-\frac{(y-\mu_Y)^2}{2\,\sigma_Y^2}\Big) \ .$$

Since $L = 2\mu_Y/\sigma_Y^2 \cdot y$, the pdf of L can be computed by the variable substitution $y = \sigma_Y^2/(2\mu_Y) \cdot l$. This results in

$$\begin{aligned} p_L(l|+1) &= \frac{\sigma_Y^2}{2\mu_Y} \cdot p_Y\big(\frac{\sigma_Y^2}{2\mu_Y}\, l\big) \\ &= \frac{1}{\sqrt{2\pi}\,\frac{2\mu_Y}{\sigma_Y}}\ \exp\Big(-\frac{(l-\frac{2\mu_Y^2}{\sigma_Y^2})^2}{2\,\big(\frac{2\mu_Y}{\sigma_Y}\big)^2}\Big) \ , \end{aligned}$$

which is again a Gaussian distribution with mean value $\mu_L = 2\mu_Y^2/\sigma_Y^2$ and variance $\sigma_L^2 = 4\mu_Y^2/\sigma_Y^2$. Since this is a conditioned pdf of an LLR, this distribution is exponential-symmetric. It is easy to prove that actually every Gaussian distribution with $\sigma^2 = 2\mu$ shows this property.

3.3. Proof of Method 3

Given Equations (8), (10) and particularly (11), Method 3 can be derived as follows:

$$\begin{aligned} P_b &= P_U(+1)\int_{l<0} p_L(l|+1)\ dl \\ &\quad + \underbrace{P_U(-1)\int_{l>0} \underbrace{p_L(l|-1)}_{\overset{(8)}{=}\, p_L(-l|+1)}\ dl}_{\int_{l<0} p_L(l|+1)} \\ &= \underbrace{[P_U(+1) + P_U(-1)]}_{1}\int_{l<0} p_L(l|+1)\ dl \\ &= \int_{l>0} p_L(-l|+1)\ dl \\ &\overset{(10)}{=} \int_{l>0} e^{-l}p_L(l|+1)\ dl \\ &\overset{(11)}{=} \int_{l>0} \frac{e^{-l}}{1+e^{-l}}p_\Lambda(l)\ dl \\ &= \int_{l>0} \frac{1}{1+e^{l}}p_\Lambda(l)\ dl \ . \end{aligned} \qquad (12)$$

Note that in the derivation of the exponential-symmetric property a uniform input distribution $P_U(u)$ was assumed, but a generalization is possible. However, Method 3 relies on two additional assumptions: the existence of a linear channel code and a symmetric channel.

REFERENCES

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes," *Proc. IEEE ICC*, pp. 1064–1070, May 1993.

[2] J. Lodge, R. Young, P. Hoeher, and J. Hagenauer, "Separable 'MAP Filters' for the Decoding of Product and Concatenated Codes," *Proc. IEEE ICC*, pp. 1740–1745, May 1993.

[3] L.R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. Inform. Theory.*, pp. 284–287, Mar. 1974.

[4] P. Robertson, P. Hoeher, and E. Villebrun, "Optimal and Sub-Optimal Maximum a Posteriori Algorithms Suitable for Turbo Decoding," *ETT*, pp. 119–125, Mar./Apr. 1997.

[5] I. Land, P. Hoeher, and U. Sorger, "On the Interpretation of the APP Algorithm as an LLR Filter," *Proc. IEEE ISIT*, June 2000.